

WO02083837

Publication Title:

METHODS FOR IDENTIFYING SMALL MOLECULES THAT BIND SPECIFIC
RNA STRUCTURAL MOTIFS

Abstract:

Abstract of WO02083837

The present invention relates to a method for screening and identifying test compounds that bind to a preselected target ribonucleic acid ("RNA"). Direct, non-competitive binding assays are advantageously used to screen bead-based libraries of compounds for those that selectively bind to a preselected target RNA. Binding of target RNA molecules to a particular test compound is detected using any physical method that measures the altered physical property of the target RNA bound to a test compound. The structure of the test compound attached to the labeled RNA is also determined. The methods used will depend, in part, on the nature of the library screened. The methods of the present invention provide a simple, sensitive assay for high-throughput screening of libraries of compounds to identify pharmaceutical leads. Data supplied from the esp@cenet database - Worldwide

Courtesy of <http://v3.espacenet.com>

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
24 October 2002 (24.10.2002)

PCT

(10) International Publication Number
WO 02/083837 A1

- (51) International Patent Classification: C12M 1/38, 1/40, C12Q 1/68
- (21) International Application Number: PCT/US02/11758
- (22) International Filing Date: 11 April 2002 (11.04.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 60/282,966 11 April 2001 (11.04.2001) US
- (71) Applicant (for all designated States except US): PTC THERAPEUTICS, INC. [US/US]; 100 Corporate Court, Middlesex Business Center, South Plainfield, NJ 07080 (US).
- (72) Inventor; and
(75) Inventor/Applicant (for US only): ALMSTEAD, Neil, G. [US/US]; 1 Crocus Drive, Holmdel, NJ 07733 (US).
- (74) Agents: CORUZZI, Laura, A. et al.; Pennie & Edmonds LLP, 1155 Avenue of the Americas, New York, NY 10036 (US).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:**
— with international search report
— before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 02/083837 A1

(54) Title: METHODS FOR IDENTIFYING SMALL MOLECULES THAT BIND SPECIFIC RNA STRUCTURAL MOTIFS

(57) **Abstract:** The present invention relates to a method for screening and identifying test compounds that bind to a preselected target ribonucleic acid ("RNA"). Direct, non-competitive binding assays are advantageously used to screen bead-based libraries of compounds for those that selectively bind to a preselected target RNA. Binding of target RNA molecules to a particular test compound is detected using any physical method that measures the altered physical property of the target RNA bound to a test compound. The structure of the test compound attached to the labeled RNA is also determined. The methods used will depend, in part, on the nature of the library screened. The methods of the present invention provide a simple, sensitive assay for high-throughput screening of libraries of compounds to identify pharmaceutical leads.

METHODS FOR IDENTIFYING SMALL MOLECULES THAT BIND SPECIFIC RNA STRUCTURAL MOTIFS

5 This application claims the benefit of U.S. Provisional Application No.
60/282,966, filed April 11, 2001, which is incorporated herein by reference in its entirety.

1. INTRODUCTION

10 The present invention relates to a method for screening and identifying test
compounds that bind to a preselected target ribonucleic acid ("RNA"). Direct, non-
competitive binding assays are advantageously used to screen bead-based libraries of
compounds for those that selectively bind to a preselected target RNA. Binding of target
RNA molecules to a particular test compound is detected using any method that measures
the altered physical property of the target RNA bound to a test compound. The methods of
15 the present invention provide a simple, sensitive assay for high-throughput screening of
libraries of compounds to identify pharmaceutical leads.

2. BACKGROUND OF THE INVENTION

Protein-nucleic acid interactions are involved in many cellular functions,
20 including transcription, RNA splicing, mRNA decay, and mRNA translation. Readily
accessible synthetic molecules that can bind with high affinity to specific sequences of
single- or double-stranded nucleic acids have the potential to interfere with these
interactions in a controllable way, making them attractive tools for molecular biology and
medicine. Successful approaches for blocking function of target nucleic acids include using
25 duplex-forming antisense oligonucleotides (Miller, 1996, Progress in Nucl. Acid Res. &
Mol. Biol. 52:261-291; Ojwang & Rando, 1999, Achieving antisense inhibition by
oligodeoxynucleotides containing N₇ modified 2'-deoxyguanosine using tumor necrosis
factor receptor type 1, METHODS: A Companion to Methods in Enzymology 18:244-251)
and peptide nucleic acids ("PNA") (Nielsen, 1999, Current Opinion in Biotechnology
30 10:71-75), which bind to nucleic acids via Watson-Crick base-pairing. Triplex-forming
anti-gene oligonucleotides can also be designed (Ping *et al.*, 1997, RNA 3:850-860;
Aggarwal *et al.*, 1996, Cancer Res. 56:5156-5164; U.S. Patent No. 5,650,316), as well as
pyrrole-imidazole polyamide oligomers (Gottesfeld *et al.*, 1997, Nature 387:202-205; White
et al., 1998, Nature 391:468-471), which are specific for the major and minor grooves of a
35 double helix, respectively.

In addition to synthetic nucleic acids (*i.e.*, antisense, ribozymes, and triplex-forming molecules), there are examples of natural products that interfere with deoxyribonucleic acid ("DNA") or RNA processes such as transcription or translation. For example, certain carbohydrate-based host cell factors, calicheamicin oligosaccharides, interfere with the sequence-specific binding of transcription factors to DNA and inhibit transcription *in vivo* (Ho *et al.*, 1994, Proc. Natl. Acad. Sci. USA 91:9203-9207; Liu *et al.*, 1996, Proc. Natl. Acad. Sci. USA 93:940-944). Certain classes of known antibiotics have been characterized and were found to interact with RNA. For example, the antibiotic thiostreptone binds tightly to a 60-mer from ribosomal RNA (Cundliffe *et al.*, 1990, in The Ribosome: Structure, Function & Evolution (Schlessinger *et al.*, eds.) American Society for Microbiology, Washington, D.C. pp. 479-490). Bacterial resistance to various antibiotics often involves methylation at specific rRNA sites (Cundliffe, 1989, Ann. Rev. Microbiol. 43:207-233). Aminoglycosidic aminocyclitol (aminoglycoside) antibiotics and peptide antibiotics are known to inhibit group I intron splicing by binding to specific regions of the RNA (von Ahsen *et al.*, 1991, Nature (London) 353:368-370). Some of these same aminoglycosides have also been found to inhibit hammerhead ribozyme function (Stage *et al.*, 1995, RNA 1:95-101). In addition, certain aminoglycosides and other protein synthesis inhibitors have been found to interact with specific bases in 16S rRNA (Woodcock *et al.*, 1991, EMBO J. 10:3099-3103). An oligonucleotide analog of the 16S rRNA has also been shown to interact with certain aminoglycosides (Purohit *et al.*, 1994, Nature 370:659-662). A molecular basis for hypersensitivity to aminoglycosides has been found to be located in a single base change in mitochondrial rRNA (Hutchin *et al.*, 1993, Nucleic Acids Res. 21:4174-4179). Aminoglycosides have also been shown to inhibit the interaction between specific structural RNA motifs and the corresponding RNA binding protein. Zapp *et al.* (Cell, 1993, 74:969-978) has demonstrated that the aminoglycosides neomycin B, lividomycin A, and tobramycin can block the binding of Rev, a viral regulatory protein required for viral gene expression, to its viral recognition element in the IIB (or RRE) region of HIV RNA. This blockage appears to be the result of competitive binding of the antibiotics directly to the RRE RNA structural motif.

Single stranded sections of RNA can fold into complex tertiary structures consisting of local motifs such as loops, bulges, pseudoknots, guanosine quartets and turns (Chastain & Tinoco, 1991, Progress in Nucleic Acid Res. & Mol. Biol. 41:131-177; Chow & Bogdan, 1997, Chemical Reviews 97:1489-1514; Rando & Hogan, 1998, Biologic activity of guanosine quartet forming oligonucleotides in "Applied Antisense Oligonucleotide Technology" Stein. & Krieg (eds) John Wiley and Sons, New York, pages 335-352). Such

structures can be critical to the activity of the nucleic acid and affect functions such as regulation of mRNA transcription, stability, or translation (Weeks & Crothers, 1993, Science 261:1574-1577). The dependence of these functions on the native three-dimensional structural motifs of single-stranded stretches of nucleic acids makes it difficult to identify or design synthetic agents that bind to these motifs using general, simple-to-use sequence-specific recognition rules for the formation of double- and triple-helical nucleic acids used in the design of antisense and ribozyme type molecules. Approaches to screening generally involve competitive assays designed to identify compounds that disrupt the interaction between a target RNA and a physiological, host cell factor(s) that had been previously identified to specifically interact with that particular target RNA. In general, such assays require the identification and characterization of the host cell factor(s) deemed to be required for the function of the target RNA. Both the target RNA and its preselected host cell binding partner are used in a competitive format to identify compounds that disrupt or interfere with the two components in the assay.

Citation or identification of any reference in Section 2 of this application is not an admission that such reference is available as prior art to the present invention.

3. SUMMARY OF THE INVENTION

The present invention relates to methods for identifying compounds that bind to preselected target elements of nucleic acids including, but not limited to, specific RNA sequences, RNA structural motifs, and/or RNA structural elements. The specific target RNA sequences, RNA structural motifs, and/or RNA structural elements are used as targets for screening small molecules and identifying those that directly bind these specific sequences, motifs, and/or structural elements. For example, methods are described in which a preselected target RNA having a detectable label is used to screen a library of test compounds, preferably under physiologic conditions. Any complexes formed between the target RNA and a member of the library are identified using methods that detect the labeled target RNA bound to a test compound. In particular, the present invention relates to methods for using a target RNA having a detectable label to screen a bead-based library of test compounds. Compounds in the bead-based library that bind to the labeled target RNA will form a bead-based detectably labeled complex, which can be separated from the unbound beads and unbound target RNA in the liquid phase by a number of physical means, including, but not limited to, flow cytometry, affinity chromatography, manual batch mode separation, suspension of beads in electric fields, and microwave of the bead-based detectably labeled complex. The detectably labeled complex can then be identified by the label on the target

RNA and removed from the uncomplexed, unlabeled test compounds in the library. The structure of the test compound complexed with the labeled RNA is then ascertained by *de novo* structure determination of the test compounds using, for example, mass spectrometry or nuclear magnetic resonance ("NMR"). The test compounds identified are useful for any purpose to which a binding reaction may be put, for example in assay methods, diagnostic procedures, cell sorting, as inhibitors of target molecule function, as probes, as sequestering agents and the like. In addition, small organic molecules which interact specifically with target RNA molecules may be useful as lead compounds for the development of therapeutic agents.

The methods described herein for the identification of compounds that directly bind to a particular preselected target RNA are well suited for high-throughput screening. The direct binding method of the invention offers advantages over drug screening systems for competitors that inhibit the formation of naturally-occurring RNA binding protein:target RNA complexes; *i.e.*, competitive assays. The direct binding method of the invention is rapid and can be set up to be readily performed, *e.g.*, by a technician, making it amenable to high throughput screening. The method of the invention also eliminates the bias inherent in the competitive drug screening systems, which require the use of a preselected host cell factor that may not have physiological relevance to the activity of the target RNA. Instead, the methods of the invention are used to identify any compound that can directly bind to specific target RNA sequences, RNA structural motifs, and/or RNA structural elements, preferably under physiologic conditions. As a result, the compounds so identified can inhibit the interaction of the target RNA with any one or more of the native host cell factors (whether known or unknown) required for activity of the RNA *in vivo*.

The present invention may be understood more fully by reference to the detailed description and examples, which are intended to illustrate non-limiting embodiments of the invention.

3.1. Definitions

As used herein, a "target nucleic acid" refers to RNA, DNA, or a chemically modified variant thereof. In a preferred embodiment, the target nucleic acid is RNA. A target nucleic acid also refers to tertiary structures of the nucleic acids, such as, but not limited to loops, bulges, pseudoknots, guanosine quartets and turns. A target nucleic acid also refers to RNA elements such as, but not limited to, the HIV TAR element, internal ribosome entry site, "slippery site", instability elements, and adenylate uridylate-rich

elements, which are described in Section 4.1. Non-limiting examples of target nucleic acids are presented in Section 4.1 and Section 5.

5 As used herein, a "library" refers to a plurality of test compounds with which a target nucleic acid molecule is contacted. A library can be a combinatorial library, *e.g.*, a collection of test compounds synthesized using combinatorial chemistry techniques, or a collection of unique chemicals of low molecular weight (less than 1000 daltons) that each occupy a unique three-dimensional space.

10 As used herein, a "label" or "detectable label" is a composition that is detectable, either directly or indirectly, by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include radioactive isotopes (*e.g.*, ^{32}P , ^{35}S , and ^3H), dyes, fluorescent dyes, electron-dense reagents, enzymes and their substrates (*e.g.*, as commonly used in enzyme-linked immunoassays, *e.g.*, alkaline phosphatase and horse radish peroxidase), biotin, digoxigenin, or haptens and proteins for
15 which antisera or monoclonal antibodies are available. Moreover, a label or detectable moiety can include an "affinity tag" that, when coupled with the target nucleic acid and incubated with a test compound or compound library, allows for the affinity capture of the target nucleic acid along with molecules bound to the target nucleic acid. One skilled in the art will appreciate that a affinity tag bound to the target nucleic acids has, by definition, a
20 complimentary ligand coupled to a solid support that allows for its capture. For example, useful affinity tags and complimentary ligands include, but are not limited to, biotin-streptavidin, complimentary nucleic acid fragments (*e.g.*, oligo dT-oligo dA, oligo T-oligo A, oligo dG-oligo dC, oligo G-oligo C), aptamer complexes, or haptens and proteins for which antisera or monoclonal antibodies are available. The label or detectable moiety is
25 typically bound, either covalently, through a linker or chemical bound, or through ionic, van der Waals or hydrogen bonds to the molecule to be detected.

As used herein, a "dye" refers to a molecule that, when exposed to radiation, emits radiation at a level that is detectable visually or via conventional spectroscopic means. As used herein, a "visible dye" refers to a molecule having a chromophore that absorbs
30 radiation in the visible region of the spectrum (*i.e.*, having a wavelength of between about 400 nm and about 700 nm) such that the transmitted radiation is in the visible region and can be detected either visually or by conventional spectroscopic means. As used herein, an "ultraviolet dye" refers to a molecule having a chromophore that absorbs radiation in the ultraviolet region of the spectrum (*i.e.*, having a wavelength of between about 30 nm and
35 about 400 nm). As used herein, an "infrared dye" refers to a molecule having a chromophore that absorbs radiation in the infrared region of the spectrum (*i.e.*, having a wavelength

between about 700 nm and about 3,000 nm). A "chromophore" is the network of atoms of the dye that, when exposed to radiation, emits radiation at a level that is detectable visually or via conventional spectroscopic means. One of skill in the art will readily appreciate that although a dye absorbs radiation in one region of the spectrum, it may emit radiation in another region of the spectrum. For example, an ultraviolet dye may emit radiation in the visible region of the spectrum. One of skill in the art will also readily appreciate that a dye can transmit radiation or can emit radiation via fluorescence or phosphorescence.

The phrase "pharmaceutically acceptable salt(s)," as used herein includes but is not limited to salts of acidic or basic groups that may be present in test compounds identified using the methods of the present invention. Test compounds that are basic in nature are capable of forming a wide variety of salts with various inorganic and organic acids. The acids that can be used to prepare pharmaceutically acceptable acid addition salts of such basic compounds are those that form non-toxic acid addition salts, *i.e.*, salts containing pharmacologically acceptable anions, including but not limited to sulfuric, citric, maleic, acetic, oxalic, hydrochloride, hydrobromide, hydroiodide, nitrate, sulfate, bisulfate, phosphate, acid phosphate, isonicotinate, acetate, lactate, salicylate, citrate, acid citrate, tartrate, oleate, tannate, pantothenate, bitartrate, ascorbate, succinate, maleate, gentisinate, fumarate, gluconate, glucuronate, saccharate, formate, benzoate, glutamate, methanesulfonate, ethanesulfonate, benzenesulfonate, p-toluenesulfonate and pamoate (*i.e.*, 1,1'-methylene-bis-(2-hydroxy-3-naphthoate)) salts. Test compounds that include an amino moiety may form pharmaceutically or cosmetically acceptable salts with various amino acids, in addition to the acids mentioned above. Test compounds that are acidic in nature are capable of forming base salts with various pharmacologically or cosmetically acceptable cations. Examples of such salts include alkali metal or alkaline earth metal salts and, particularly, calcium, magnesium, sodium lithium, zinc, potassium, and iron salts.

By "substantially one type of test compound," as used herein, is meant that the assay can be performed in such a fashion that at some point, only one compound need be used in each reaction so that, if the result is indicative of a binding event occurring between the target RNA molecule and the test compound the test compound, can be easily identified.

4. DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to methods for identifying compounds that bind to preselected target elements of nucleic acids, in particular, RNAs, including but not limited to preselected target RNA sequencing structural motifs, or structural elements. Methods are described in which a preselected target RNA having a detectable label is used to screen a

library of test compounds. Any complexes formed between the target RNA and a member of the library are identified using methods that detect the labeled target RNA bound to a test compound. In particular, the present invention relates to methods for using a target RNA having a detectable label to screen a bead-based library of test compounds. Compounds in the bead-based library that bind to the labeled target RNA will form a bead-based detectably labeled complex, which can be separated from the unbound target RNA in the liquid phase by a number of physical means, such as, but not limited to, flow cytometry, affinity chromatography, manual batch mode separation, suspension of beads in electric fields, and microwave of the bead-based detectably labeled complex. The detectably labeled complex can then be identified by the label on the target RNA and removed from the uncomplexed, unlabeled test compounds in the library. The structure of the test compound attached to the labeled RNA is then ascertained by *de novo* structure determination of the test compounds using, for example, mass spectrometry or nuclear magnetic resonance ("NMR").

Thus, the methods of the present invention provide a simple, sensitive assay for high-throughput screening of libraries of test compounds, in which the test compounds of the library that specifically bind a preselected target nucleic acid are easily distinguished from non-binding members of the library. The structures of the binding molecules are ascertained by *de novo* structure determination of the test compounds using, for example, mass spectrometry or nuclear magnetic resonance ("NMR"). The test compounds so identified are useful for any purpose to which a binding reaction may be put, for example in assay methods, diagnostic procedures, cell sorting, as inhibitors of target molecule function, as probes, as sequestering agents and lead compounds for development of therapeutics, and the like. Small organic compounds that are identified to interact specifically with the target RNA molecules are particularly attractive candidates as lead compounds for the development of therapeutic agents.

The assay of the invention reduces bias introduced by competitive binding assays which require the identification and use of a host cell factor (presumably essential for modulating RNA function) as a binding partner for the target RNA. The assays of the present invention are designed to detect any compound or agent that binds to the target RNA, preferably under physiologic conditions. Such agents can then be tested for biological activity, without establishing or guessing which host cell factor or factors is required for modulating the function and/or activity of the target RNA.

Section 4.1 describes examples of protein-RNA interactions that are important in a variety of cellular functions and several target RNA elements that can be used to identify test compounds. Compounds that inhibit these interactions by binding to the RNA and

successfully competing with the natural protein or host cell factor that endogenously binds to the RNA may be important, *e.g.*, in treating or preventing a disease or abnormal condition, such as an infection or unchecked growth. Section 4.2 describes detectable labels for target nucleic acids that are useful in the methods of the invention. Section 4.3 describes libraries of test compounds. Section 4.4 provides conditions for binding a labeled target RNA to a test compound of a library and detecting RNA binding to a test compound using the methods of the invention. Section 4.5 provides methods for separating complexes of target RNAs bound to a test compound from an unbound RNA. Section 4.6 describes methods for identifying test compounds that are bound to the target RNA. Section 4.7 describes a secondary, biological screen of test compounds identified by the methods of the invention to test the effect of the test compounds *in vivo*. Section 4.8 describes the use of test compounds identified by the methods of the invention for treating or preventing a disease or abnormal condition in mammals.

4.1. Biologically Important RNA-Host Cell Factor Interactions

Nucleic acids, and in particular RNAs, are capable of folding into complex tertiary structures that include bulges, loops, triple helices and pseudoknots, which can provide binding sites for host cell factors, such as proteins and other RNAs. RNA-protein and RNA-RNA interactions are important in a variety cellular functions, including transcription, RNA splicing, RNA stability and translation. Furthermore, the binding of such host cell factors to RNAs may alter the stability and translational efficiency of such RNAs, and according affect subsequent translation. For example, some diseases are associated with protein overproduction or decreased protein function. In this case, the identification of compounds to modulate RNA stability and translational efficiency will be useful to treat and prevent such diseases.

The methods of the present invention are useful for identifying test compounds that bind to target RNA elements in a high throughput screening assay of libraries of test compounds in solution. In particular, the methods of the present invention are useful for identifying a test compound that binds to a target RNA elements and inhibits the interaction of that RNA with one or more host cell factors *in vivo*. The molecules identified using the methods of the invention are useful for inhibiting the formation of a specific bound RNA:host cell factor complexes *in vivo*.

In some embodiments, test compounds identified by the methods of the invention are useful for increasing or decreasing the translation of messenger RNAs ("mRNAs"), *e.g.*, protein production, by binding to one or more regulatory elements in the 5'

untranslated region, the 3' untranslated region, or the coding region of the mRNA.

Compounds that bind to mRNA can, *inter alia*, increase or decrease the rate of mRNA processing, alter its transport through the cell, prevent or enhance binding of the mRNA to ribosomes, suppressor proteins or enhancer proteins, or alter mRNA stability. Accordingly, compounds that increase or decrease mRNA translation can be used to treat or prevent disease. For example, diseases associated with protein overproduction, such as amyloidosis, or with the production of mutant proteins, such as *Ras*, can be treated or prevented by decreasing translation of the mRNA that codes for the overproduced protein, thus inhibiting production of the protein. Conversely, the symptoms of diseases associated with decreased protein function, such as hemophilia, may be treated by increasing translation of mRNA coding for the protein whose function is decreased, *e.g.*, factor IX in some forms of hemophilia.

The methods of the invention can be used to identify compounds that bind to mRNAs coding for a variety of proteins with which the progression of diseases in mammals is associated. These mRNAs include, but are not limited to, those coding for amyloid protein and amyloid precursor protein; anti-angiogenic proteins such as angiostatin, endostatin, METH-1 and METH-2; apoptosis inhibitor proteins such as survivin, clotting factors such as Factor IX, Factor VIII, and others in the clotting cascade; collagens; cyclins and cyclin inhibitors, such as cyclin dependent kinases, cyclin D1, cyclin E, WAF1, cdk4 inhibitor, and MTS1; cystic fibrosis transmembrane conductance regulator gene (CFTR); cytokines such as IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IL-16, IL-17 and other interleukins; hematopoietic growth factors such as erythropoietin (Epo); colony stimulating factors such as G-CSF, GM-CSF, M-CSF, SCF and thrombopoietin; growth factors such as BDNF, BMP, GGRP, EGF, FGF, GDNF, GGF, HGF, IGF-1, IGF-2, KGF, myotrophin, NGF, OSM, PDGF, somatotrophin, TGF- β , TGF- α and VEGF; antiviral cytokines such as interferons, antiviral proteins induced by interferons, TNF- α , and TNF- β ; enzymes such as cathepsin K, cytochrome P-450 and other cytochromes, farnesyl transferase, glutathione-s transferases, heparanase, HMG CoA synthetase, N-acetyltransferase, phenylalanine hydroxylase, phosphodiesterase, ras carboxyl-terminal protease, telomerase and TNF converting enzyme; glycoproteins such as cadherins, *e.g.*, N-cadherin and E-cadherin; cell adhesion molecules; selectins; transmembrane glycoproteins such as CD40; heat shock proteins; hormones such as 5- α reductase, atrial natriuretic factor, calcitonin, corticotrophin releasing factor, diuretic hormones, glucagon, gonadotropin, gonadotropin releasing hormone, growth hormone, growth hormone releasing factor, somatotropin, insulin, leptin, luteinizing hormone, luteinizing hormone releasing hormone,

parathyroid hormone, thyroid hormone, and thyroid stimulating hormone; proteins involved in immune responses, including antibodies, CTLA4, hemagglutinin, MHC proteins, VLA-4, and kallikrein-kininogen-kinin system; ligands such as CD4; oncogene products such as *sis*, *hst*, protein tyrosine kinase receptors, *ras*, *abl*, *mos*, *myc*, *fos*, *jun*, *H-ras*, *ki-ras*, *c-fms*, *bcl-2*, *L-myc*, *c-myc*, *gip*, *gsp*, and *HER-2*; receptors such as bombesin receptor, estrogen receptor, GABA receptors, growth factor receptors including EGFR, PDGFR, FGFR, and NGFR, GTP-binding regulatory proteins, interleukin receptors, ion channel receptors, leukotriene receptor antagonists, lipoprotein receptors, opioid pain receptors, substance P receptors, retinoic acid and retinoid receptors, steroid receptors, T-cell receptors, thyroid hormone receptors, TNF receptors; tissue plasminogen activator; transmembrane receptors; transmembrane transporting systems, such as calcium pump, proton pump, Na/Ca exchanger, MRP1, MRP2, P170, LRP, and cMOAT; transferrin; and tumor suppressor gene products such as *APC*, *brca1*, *brca2*, *DCC*, *MCC*, *MTS1*, *NF1*, *NF2*, *nm23*, *p53* and *Rb*. In addition to the eukaryotic genes listed above, the invention, as described, can be used to define molecules that interrupt viral, bacterial or fungal transcription or translation efficiencies and therefore form the basis for a novel anti-infectious disease therapeutic. Other target genes include, but are not limited to, those disclosed in Section 4.1 and Section 5.

The methods of the invention can be used to identify mRNA-binding test compounds for increasing or decreasing the production of a protein, thus treating or preventing a disease associated with decreasing or increasing the production of said protein, respectively. The methods of the invention may be useful for identifying test compounds for treating or preventing a disease in mammals, including cats, dogs, swine, horses, goats, sheep, cattle, primates and humans. Such diseases include, but are not limited to, amyloidosis, hemophilia, Alzheimer's disease, atherosclerosis, cancer, gigantism, dwarfism, hypothyroidism, hyperthyroidism, inflammation, cystic fibrosis, autoimmune disorders, diabetes, aging, obesity, neurodegenerative disorders, and Parkinson's disease. Other diseases include, but are not limited to, those described in Section 4.1 and diseases caused by aberrant expression of the genes disclosed in Example 5. In addition to the eukaryotic genes listed above, the invention, as described, can be used to define molecules that interrupt viral, bacterial or fungal transcription or translation efficiencies and therefore form the basis for a novel anti-infectious disease therapeutic.

In other embodiments, test compounds identified by the methods of the invention are useful for preventing the interaction of an RNA, such as a transfer RNA ("tRNA"), an enzymatic RNA or a ribosomal RNA ("rRNA"), with a protein or with another RNA, thus preventing, e.g., assembly of an *in vivo* protein-RNA or RNA-RNA complex that

is essential for the viability of a cell. The term "enzymatic RNA," as used herein, refers to RNA molecules that are either self-splicing, or that form an enzyme by virtue of their association with one or more proteins, *e.g.*, as in RNase P, telomerase or small nuclear ribonuclear protein particles. For example, inhibition of an interaction between rRNA and one or more ribosomal proteins may inhibit the assembly of ribosomes, rendering a cell incapable of synthesizing proteins. In addition, inhibition of the interaction of precursor rRNA with ribonucleases or ribonucleoprotein complexes (such as RNase P) that process the precursor rRNA prevent maturation of the rRNA and its assembly into ribosomes. Similarly, a tRNA:tRNA synthetase complex may be inhibited by test compounds identified by the methods of the invention such that tRNA molecules do not become charged with amino acids. Such interactions include, but are not limited to, rRNA interactions with ribosomal proteins, tRNA interactions with tRNA synthetase, RNase P protein interactions with RNase P RNA, and telomerase protein interactions with telomerase RNA.

In other embodiments, test compounds identified by the methods of the invention are useful for treating or preventing a viral, bacterial, protozoan or fungal infection. For example, transcriptional up-regulation of the genes of human immunodeficiency virus type 1 ("HIV-1") requires binding of the HIV Tat protein to the HIV trans-activation response region RNA ("TAR RNA"). HIV TAR RNA is a 59-base stem-loop structure located at the 5'-end of all nascent HIV-1 transcripts (Jones & Peterlin, 1994, *Annu. Rev. Biochem.* 63:717-43). Tat protein is known to interact with uracil 23 in the bulge region of the stem of TAR RNA. Thus, TAR RNA is a potential binding target for test compounds, such as small peptides and peptide analogs that bind to the bulge region of TAR RNA and inhibit formation of a Tat-TAR RNA complex involved in HIV-1 upregulation (see Hwang *et al.*, 1999 *Proc. Natl. Acad. Sci. USA* 96:12997-13002). Accordingly, test compounds that bind to TAR RNA are useful as anti-HIV therapeutics (Hamy *et al.*, 1997, *Proc. Natl. Acad. Sci. USA* 94:3548-3553; Hamy *et al.*, 1998, *Biochemistry* 37:5086-5095; Mei *et al.*, 1998, *Biochemistry* 37:14204-14212), and therefore, are useful for treating or preventing AIDS.

The methods of the invention can be used to identify test compounds to treat or prevent viral, bacterial, protozoan or fungal infections in a patient. In some embodiments, the methods of the invention are useful for identifying compounds that decrease translation of microbial genes by interacting with mRNA, as described above, or for identifying compounds that inhibit the interactions of microbial RNAs with proteins or other ligands that are essential for viability of the virus or microbe. Examples of microbial target RNAs useful in the present invention for identifying antiviral, antibacterial, anti-protozoan and anti-fungal compounds include, but are not limited to, general antiviral and anti-inflammatory targets

such as mRNAs of INF α , INF γ , RNase L, RNase L inhibitor protein, PKR, tumor necrosis factor, interleukins 1-15, and IMP dehydrogenase; internal ribosome entry sites; HIV-1 CT rich domain and RNase H mRNA; HCV internal ribosome entry site (required to direct translation of HCV mRNA), and the 3'-untranslated tail of HCV genomes; rotavirus NSP3 binding site, which binds the protein NSP3 that is required for rotavirus mRNA translation; HBV epsilon domain; Dengue virus 5' and 3' untranslated regions, including IRES; INF α , INF β and INF γ ; plasmodium falciparum mRNAs; the 16S ribosomal subunit ribosomal RNA and the RNA component of RNase P of bacteria; and the RNA component of telomerase in fungi and cancer cells. Other target viral and bacterial mRNAs include, but are not limited to, those disclosed in Section 5.

One of skill in the art will appreciate that, although such target RNAs are functionally conserved in various species (*e.g.*, from yeast to humans), they exhibit nucleotide sequence and structural diversity. Therefore, inhibition of, for example, yeast telomerase by an anti-fungal compound identified by the methods of the invention might not interfere with human telomerase and normal human cell proliferation.

Thus, the methods of the invention can be used to identify test compounds that interfere with one or more target RNA interactions with host cell factors that are important for cell growth or viability, or essential in the life cycle of a virus, a bacterium, a protozoa or a fungus. Such test compounds and/or congeners that demonstrate desirable biologic and pharmacologic activity can be administered to a patient in need thereof in order to treat or prevent a disease caused by viral, bacterial, protozoan, or fungal infections. Such diseases include, but are not limited to, HIV infection, AIDS, human T-cell leukemia, SIV infection, FIV infection, feline leukemia, hepatitis A, hepatitis B, hepatitis C, Dengue fever, malaria, rotavirus infection, severe acute gastroenteritis, diarrhea, encephalitis, hemorrhagic fever, syphilis, legionella, whooping cough, gonorrhea, sepsis, influenza, pneumonia, tinea infection, candida infection, and meningitis.

Non-limiting examples of RNA elements involved in the regulation of gene expression, *i.e.*, mRNA stability, translational efficiency via translational initiation and ribosome assembly, *etc.*, include the HIV TAR element, internal ribosome entry site, "slippery site", instability elements, and adenylate uridylate-rich elements, as discussed below.

4.1.1. HIV TAR Element

Transcriptional up-regulation of the genes of human immunodeficiency virus type 1 ("HIV-1") requires binding of the HIV Tat protein to the HIV trans-activation

response region RNA ("TAR RNA"), a 59-base stem-loop structure located at the 5' end of all nascent HIV-1 transcripts (Jones & Peterlin, 1994, *Annu. Rev. Biochem.* 63:717-43). Tat protein is known to interact with uracil 23 in the bulge region of the stem of TAR RNA.

Thus, TAR RNA is a useful binding target for test compounds, such as small peptides and peptide analogs that bind to the bulge region of TAR RNA and inhibit formation of a Tat-TAR RNA complex involved in HIV-1 up-regulation (see Hwang *et al.*, 1999 *Proc. Natl. Acad. Sci. USA* 96:12997-13002). Accordingly, test compounds that bind to TAR RNA can be useful as anti-HIV therapeutics (Hamy *et al.*, 1997, *Proc. Natl. Acad. Sci. USA* 94:3548-3553; Hamy *et al.*, 1998, *Biochemistry* 37:5086-5095; Mei *et al.*, 1998, *Biochemistry* 37:14204-14212), and therefore, are useful for treating or preventing AIDS.

4.1.2. Internal Ribosome Entry Site ("IRES")

Internal ribosome entry sites ("IRES") are found in the 5' untranslated regions ("5' UTR") of several mRNAs, and are thought to be involved in the regulation of translational efficiency. When the IRES element is present on an mRNA downstream of a translational stop codon, it directs ribosomal re-entry (Ghattas *et al.*, 1991, *Mol. Cell. Biol.* 11:5848-5959), which permits initiation of translation at the start of a second open reading frame.

As reviewed by Jang *et al.*, a large segment of the 5' nontranslated region, approximately 400 nucleotides in length, promotes internal entry of ribosomes independent of the non-capped 5' end of picornavirus mRNAs (mammalian plus-strand RNA viruses whose genomes serve as mRNA). This 400 nucleotide segment (IRES), maps approximately 200 nt down-stream from the 5' end and is highly structured. IRES elements of different picornaviruses, although functionally similar *in vitro* and *in vivo*, are not identical in sequence or structure. However, IRES elements of the genera enteroviruses, on the one hand, and cardioviruses, on the other hand, reveal similarities corresponding to phylogenetic kinship. All IRES elements contain a conserved Yn-Xm-AUG unit (Y, pyrimidine; X, nucleotide) which appears essential for IRES function. The IRES elements of cardioviruses, enteroviruses and aphthoviruses bind a cellular protein, p57. In the case of cardioviruses, the interaction between a specific stem-loop of the IREs is essential for translation *in vitro*. The IRES elements of enteroviruses and cardioviruses also bind the cellular protein, p52, but the significance of this interaction remains to be shown. The function of p57 or p52 in cellular metabolism is unknown. Since picornaviral IRES elements function *in vivo* in the absence of any viral gene products, it is speculated that IRES-like elements may also occur in specific cellular mRNAs releasing them from cap-dependent translation (Jang *et al.*,

1990, Enzyme 44(1-4):292-309).

4.1.3. "Slippery Site"

5 Programmed, or directed, ribosomal frameshifting, when ribosomes shift from one translation reading frame to another and synthesize two viral proteins from a single viral mRNA, is directed by a unique site in viral mRNAs called the "slippery site." The slippery site directs ribosomal frameshifting in the -1 or +1 direction that causes the ribosome to slip by one base in the 5' direction thereby placing the ribosome in the new reading frame to produce a new protein.

10 Programmed, or directed, ribosomal frameshifting is of particular value to viruses that package their plus strands, as it eliminates the need to splice their mRNAs and reduces the risk of packaging defective genomes and regulates the ratio of viral proteins synthesized. Examples of programmed translational frameshifting (both +1 and -1 shifts) have been identified in ScV systems (Lopinski *et al.*, 2000, Mol. Cell. Biol. 20(4):1095-103, 15 retroviruses (Falk *et al.*, 1993, J. Virol. 67:273-6277; Jacks & Varmus, 1985, Science 230:1237-1242; Morikawa & Bishop, 1992, Virology 186:389-397; Nam *et al.*, 1993, J. Virol. 67:196-203); coronaviruses (Brierley *et al.*, 1987, EMBO J. 6:3779-3785; Herold & Siddell, 1993, Nucleic Acids Res. 21:5838-5842); giardiaviruses, which are also members of the Totiviridae (Wang *et al.*, 1993, Proc. Natl. Acad. Sci. USA 90:8595-8599); two bacterial 20 genes (Blinkowa & Walker, 1990, Nucleic Acids Res., 18:1725-1729; Craig & Caskey, 1986, Nature 322:273); bacteriophage genes (Condron *et al.*, 1991, Nucleic Acids Res. 19:5607-5612); astroviruses (Marczinke *et al.*, 1994, J. Virol. 68:5588-5595); the yeast EST3 gene (Lundblad & Morris, 1997, Curr. Biol. 7:969-976); and the rat, mouse, *Xenopus*, and 25 *Drosophila* ornithine decarboxylase antizymes (Matsufuji *et al.*, 1995, Cell 80:51-60); and a significant number of cellular genes (Herold & Siddell, 1993, Nucleic Acids Res. 21:5838-5842).

Drugs targeted to ribosomal frameshifting minimize the problem of virus drug resistance because this strategy targets a host cellular process rather than one introduced into the cell by the virus, which minimizes the ability of viruses to evolve drug-resistant mutants. 30 Compounds that target the RNA elements involved in regulating programmed frameshifting should have several advantages, including (a) any selective pressure on the host cellular translational machinery to adapt to the drugs would have to occur at the host evolutionary time scale, which is on the order of millions of years, (b) ribosomal frameshifting is not used to express any host proteins, and (c) altering viral frameshifting efficiencies by modulating 35

the activity of a host protein minimizing the likelihood that the virus will acquire resistance to such inhibition by mutations in its own genome.

4.1.4. Instability Elements

“Instability elements” may be defined as specific sequence elements that promote the recognition of unstable mRNAs by cellular turnover machinery. Instability elements have been found within mRNA protein coding regions as well as untranslated regions.

Altering the control of stability of normal mRNAs may lead to disease. The alteration of mRNA stability has been implicated in diseases such as, but not limited to, cancer, immune disorders, heart disease, and fibrotic disorders.

There are several examples of mutations that delete instability elements which then result in stabilization of mRNAs that may be involved in the onset of cancer. In Burkitt's lymphoma, a portion of the *c-myc* proto-oncogene is translocated to an Ig locus, producing a form of the *c-myc* mRNA that is five times more stable (*see, e.g.,* Kapstein *et al.,* 1996, *J. Biol. Chem.* 271(31):18875-84). The highly oncogenic *v-fos* mRNA lacks the 3' UTR adenylate uridylate rich element (“ARE”) that is found in the more labile and weakly oncogenic *c-fos* mRNA (*see, e.g.,* Schiavi *et al.,* 1992, *Biochim Biophys Acta.* 1114(2-3):95-106). Differences between the benign cervical lesions brought about by nonintegrated circular human papillomavirus type 16 and its integrated form, that lacks the 3' UTR ARE and correlates with cervical carcinomas, may be a consequence of stabilizing the E6/E7 transcripts encoding oncogenic proteins. Integration of the virus results in deletion of the ARE instability element, resulting in stabilization of the transcripts and over-expression of the proteins (*see, e.g.,* Jeon & Lambert, 1995, *Proc. Natl. Acad. Sci. USA* 92(5):1654-8). Deletion of AREs from the 3' UTR of the IL-2 and IL-3 genes promotes increased stabilization of these mRNAs, high expression of these proteins, and leads to the formation of cancerous cells (*see, e.g.,* Stoecklin *et al.,* 2000, *Mol. Cell. Biol.* 20(11):3753-63).

Mutations in trans-acting factors involved in mRNA turnover may also promote cancer. In monocytic tumors, the lymphokine GM-CSF mRNA is specifically stabilized as a consequence of an oncogenic lesion in a trans-acting factor that controls mRNA turnover rates. Furthermore, the normally unstable IL-3 transcript is inappropriately long-lived in mast tumor cells. Similarly, the labile GM-CSF mRNA is greatly stabilized in bladder carcinoma cells. *See, e.g.,* Bickel *et al.,* 1990, *J. Immunol.* 145(3):840-5.

The immune system is regulated by a large number of regulatory molecules that either activate or inhibit the immune response. It has now been clearly demonstrated that

stability of the transcripts encoding these proteins are highly regulated. Altered regulation of these molecules leads to mis-regulation of this process and can result in drastic medical consequences. For example, recent results using transgenic mice have shown that mis-regulation of the stability of the important modulator TNF α mRNA leads to diseases such as, but not limited to, rheumatoid arthritis and a Crohn's-like liver disease. *See, e.g., Clark, 2000, Arthritis Res. 2(3):172-4.*

Smooth muscle in the heart is modulated by the β -adrenergic receptor, which in turn responds to the sympathetic neurotransmitter norepinephrine and the adrenal hormone epinephrine. Chronic heart failure is characterized by impairment of smooth muscle cells, which results, in part, from the more rapid decay of the β -adrenergic receptor mRNA. *See, e.g., Ellis & Frielle T., 1999, Biochem. Biophys. Res. Commun. 258(3):552-8.*

A large number of diseases result from over-expression of collagen. For example, cirrhosis results from damage to the liver as a consequence of cancer, viral infection, or alcohol abuse. Such damage causes mis-regulation of collagen expression, leading to the formation of large collagen deposits. Recent results indicate that the sizeable increase in collagen expression is largely attributable to stabilization of its mRNA. *See, e.g., Lindquist et al., 2000, Am. J. Physiol. Gastrointest. Liver Physiol. 279(3):G471-6.*

4.1.5. Adenylate Uridylate-rich Elements ("ARE")

Adenylate uridylate-rich elements ("ARE") are found in the 3' untranslated regions ("3' UTR") of several mRNAs, and involved in the turnover of mRNAs, such as but not limited to transcription factors, cytokines, and lymphokines. AREs may function both as stabilizing and destabilizing elements. ARE mRNAs are classified into five groups, depending on sequence (Bakheet *et al.*, 2001, Nucl. Acids Res. 29(1):246-254). An ongoing database at the web site <http://rc.kfshrc.edu.sa/ared> contains ARE-containing mRNAs and their cluster groups, which is incorporated by reference in its entirety. The ARE motifs are classified as follows:

Group I Cluster	(AUUUUUUUUUUUUUUUUUUU)	SEQ ID NO: 1
Group II Cluster	(AUUUUUUUUUUUUUUUUU) stretch	SEQ ID NO: 2
Group III Cluster	(WAUUUUUUUUUUUUUUUU) stretch	SEQ ID NO: 3
Group IV Cluster	(WWAUUUUUUUUUUUUUUU) stretch	SEQ ID NO: 4
Group V Cluster	(WWWWAUUUUUUUUUUUUUUU) stretch	SEQ ID NO: 5

The ARE-mRNAs were clustered into five groups containing five, four, three and two pentameric repeats, while the last group contains only one pentamer within the

13-bp ARE pattern. Functional categories were assigned whenever possible according to NCBI-COG functional annotation (Tatusov *et al.*, 2001, Nucleic Acids Research, 29(1): 22-28), in addition to the categories: inflammation, immune response, development/differentiation, using an extensive literature search.

Group I contains many secreted proteins including GM-CSF, IL-1, IL-11, IL-12 and Gro- β that affect the growth of hematopoietic and immune cells (Witsell & Schook, 1992, Proc. Natl Acad. Sci. USA, 89:4754-4758). Although TNF α is both a pro-inflammatory and anti-tumor protein, there is experimental evidence that it can act as a growth factor in certain leukemias and lymphomas (Liu *et al.*, 2000, J. Biol. Chem. 275:21086-21093).

Unlike Group I, Groups II-V contain functionally diverse gene families comprising immune response, cell cycle and proliferation, inflammation and coagulation, angiogenesis, metabolism, energy, DNA binding and transcription, nutrient transportation and ionic homeostasis, protein synthesis, cellular biogenesis, signal transduction, and apoptosis (Bakheet *et al.*, 2001, Nucl. Acids Res. 29(1):246-254).

Several groups have described ARE-binding proteins that influence the ARE-mRNA stability. Among the well-characterized proteins are the mammalian homologs of ELAV (embryonic lethal abnormal vision) proteins including AUF1, HuR and He1-N2 (Zhang *et al.*, 1993, Mol. Cell. Biol. 13:7652-7665; Levine *et al.*, 1993, Mol. Cell. Biol. 13:3494-3504; Ma *et al.*, 1996, J. Biol. Chem. 271:8144-8151). The zinc-finger protein tristetraprolin has been identified as another ARE-binding protein with destabilizing activity on TNF α , IL-3 and GM-CSF mRNAs (Stoecklin *et al.*, 2000, Mol. Cell. Biol. 20:3753-3763; Carballo *et al.*, 2000, Blood 95:1891-1899).

Since ARE-containing genes are clearly important in biological systems, including but not limited to a number of the early response genes that regulate cell proliferation and responses to exogenous agents, the identification of compounds that bind to one or more of the ARE clusters and potentially modulate the stability of the target RNA can potentially be of value as a therapeutic.

4.2. Detectably Labeled Target RNAs

Target nucleic acids, including but not limited to RNA and DNA, useful in the methods of the present invention have a label that is detectable via conventional spectroscopic means or radiographic means. Preferably, target nucleic acids are labeled with a covalently attached dye molecule. Useful dye-molecule labels include, but are not limited

to, fluorescent dyes, phosphorescent dyes, ultraviolet dyes, infrared dyes, and visible dyes. Preferably, the dye is a visible dye.

Useful labels in the present invention can include, but are not limited to, spectroscopic labels such as fluorescent dyes (*e.g.*, fluorescein and derivatives such as fluorescein isothiocyanate (FITC) and Oregon Green™, rhodamine and derivatives (*e.g.*, Texas red, tetramethylrhodimine isothiocyanate (TRITC), bora-3a,4a-diaza-s-indacene (BODIPY®) and derivatives, *etc.*), digoxigenin, biotin, phycoerythrin, AMCA, CyDye™, and the like), radiolabels (*e.g.*, ³H, ¹²⁵I, ³⁵S, ¹⁴C, ³²P, ³³P, *etc.*), enzymes (*e.g.*, horse radish peroxidase, alkaline phosphatase *etc.*), spectroscopic colorimetric labels such as colloidal gold or colored glass or plastic (*e.g.* polystyrene, polypropylene, latex, *etc.*) beads, or nanoparticles – nanoclusters of inorganic ions with defined dimension from 0.1 to 1000 nm. The label may be coupled directly or indirectly to a component of the detection assay (*e.g.*, the detection reagent) according to methods well known in the art. A wide variety of labels may be used, with the choice of label depending on sensitivity required, ease of conjugation with the compound, stability requirements, available instrumentation, and disposal provisions.

In one embodiment, nucleic acids that are labeled at one or more specific locations are chemically synthesized using phosphoramidite or other solution or solid-phase methods. Detailed descriptions of the chemistry used to form polynucleotides by the phosphoramidite method are well known (*see, e.g.*, Caruthers *et al.*, U.S. Pat. Nos. 4,458,066 and 4,415,732; Caruthers *et al.*, 1982, Genetic Engineering 4:1-17; *Users Manual Model 392 and 394 Polynucleotide Synthesizers*, 1990, pages 6-1 through 6-22, Applied Biosystems, Part No. 901237; Ojwang, *et al.*, 1997, Biochemistry, 36:6033-6045). The phosphoramidite method of polynucleotide synthesis is the preferred method because of its efficient and rapid coupling and the stability of the starting materials. The synthesis is performed with the growing polynucleotide chain attached to a solid support, such that excess reagents, which are generally in the liquid phase, can be easily removed by washing, decanting, and/or filtration, thereby eliminating the need for purification steps between synthesis cycles.

The following briefly describes illustrative steps of a typical polynucleotide synthesis cycle using the phosphoramidite method. First, a solid support to which is attached a protected nucleoside monomer at its 3' terminus is treated with acid, *e.g.*, trichloroacetic acid, to remove the 5'-hydroxyl protecting group, freeing the hydroxyl group for a subsequent coupling reaction. After the coupling reaction is completed an activated intermediate is formed by contacting the support-bound nucleoside with a protected nucleoside phosphoramidite monomer and a weak acid, *e.g.*, tetrazole. The weak acid

protonates the nitrogen atom of the phosphoramidite forming a reactive intermediate. Nucleoside addition is generally complete within 30 seconds. Next, a capping step is performed, which terminates any polynucleotide chains that did not undergo nucleoside addition. Capping is preferably performed using acetic anhydride and 1-methylimidazole. The phosphite group of the internucleotide linkage is then converted to the more stable phosphotriester by oxidation using iodine as the preferred oxidizing agent and water as the oxygen donor. After oxidation, the hydroxyl protecting group of the newly added nucleoside is removed with a protic acid, e.g., trichloroacetic acid or dichloroacetic acid, and the cycle is repeated one or more times until chain elongation is complete. After synthesis, the polynucleotide chain is cleaved from the support using a base, e.g., ammonium hydroxide or t-butyl amine. The cleavage reaction also removes any phosphate protecting groups, e.g., cyanoethyl. Finally, the protecting groups on the exocyclic amines of the bases and any protecting groups on the dyes are removed by treating the polynucleotide solution in base at an elevated temperature, e.g., at about 55°C. Preferably the various protecting groups are removed using ammonium hydroxide or t-butyl amine.

Any of the nucleoside phosphoramidite monomers can be labeled using standard phosphoramidite chemistry methods (Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96(23):12997-13002; Ojwang *et al.*, 1997, Biochemistry. 36:6033-6045 and references cited therein). Dye molecules useful for covalently coupling to phosphoramidites preferably comprise a primary hydroxyl group that is not part of the dye's chromophore. Illustrative dye molecules include, but are not limited to, disperse dye CAS 4439-31-0, disperse dye CAS 6054-58-6, disperse dye CAS 4392-69-2 (Sigma-Aldrich, St. Louis, MO), disperse red, and 1-pyrenebutanol (Molecular Probes, Eugene, OR). Other dyes useful for coupling to phosphoramidites will be apparent to those of skill in the art, such as fluorescein, cy3, and cy5 fluorescent dyes, and may be purchased from, e.g., Sigma-Aldrich, St. Louis, MO or Molecular Probes, Inc., Eugene, OR.

In another embodiment, dye-labeled target RNA molecules are synthesized enzymatically using *in vitro* transcription (Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96(23):12997-13002 and references cited therein). In this embodiment, a template DNA is denatured by heating to about 90°C and an oligonucleotide primer is annealed to the template DNA, for example by slow-cooling the mixture of the denatured template and the primer from about 90°C to room temperature. A mixture of ribonucleoside-5'-triphosphates capable of supporting template-directed enzymatic extension of the primed template (e.g., a mixture including GTP, ATP, CTP, and UTP), including one or more dye-labeled ribonucleotides (Sigma-Aldrich, St. Louis, MO), is added to the primed template. Next, a polymerase

enzyme is added to the mixture under conditions where the polymerase enzyme is active, which are well-known to those skilled in the art. A labeled polynucleotide is formed by the incorporation of the labeled ribonucleotides during polymerase-mediated strand synthesis.

5 In yet another embodiment of the invention, nucleic acid molecules are end-labeled after their synthesis. Methods for labeling the 5'-end of an oligonucleotide include but are by no means limited to: (i) periodate oxidation of a 5'-to-5'-coupled ribonucleotide, followed by reaction with an amine-reactive label (Heller & Morisson, 1985, in *Rapid Detection and Identification of Infectious Agents*, D.T. Kingsbury and S. Falkow, eds., pp. 245-256, Academic Press); (ii) condensation of ethylenediamine with 5'-phosphorylated
10 polynucleotide, followed by reaction with an amine reactive label (Morrison, European Patent Application 232 967); (iii) introduction of an aliphatic amine substituent using an aminohexyl phosphite reagent in solid-phase DNA synthesis, followed by reaction with an amine reactive label (Cardullo *et al.*, 1988, Proc. Natl. Acad. Sci. USA 85:8790-8794); and
15 (iv) introduction of a thiophosphate group on the 5'-end of the nucleic acid, using phosphatase treatment followed by end-labeling with ATP- S and kinase, which reacts specifically and efficiently with maleimide-labeled fluorescent dyes (Czworkowski *et al.*, 1991, Biochem. 30:4821-4830).

A detectable label should not be incorporated into a target nucleic acid at the
20 specific binding site at which test compounds are likely to bind, since the presence of a covalently attached label might interfere sterically or chemically with the binding of the test compounds at this site. Accordingly, if the region of the target nucleic acid that binds to a host cell factor is known, a detectable label is preferably incorporated into the nucleic acid molecule at one or more positions that are spatially or sequentially remote from the binding
25 region.

After synthesis, the labeled target nucleic acid can be purified using standard techniques known to those skilled in the art (*see* Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96(23):12997-13002 and references cited therein). Depending on the length of the target nucleic acid and the method of its synthesis, such purification techniques include, but
30 are not limited to, reverse-phase high-performance liquid chromatography ("reverse-phase HPLC"), fast performance liquid chromatography ("FPLC"), and gel purification. After purification, the target RNA is refolded into its native conformation, preferably by heating to approximately 85-95°C and slowly cooling to room temperature in a buffer, *e.g.*, a buffer comprising about 50 mM Tris-HCl, pH 8 and 100 mM NaCl.

35 In another embodiment, the target nucleic acid can also be radiolabeled. A radiolabel, such as, but not limited to, an isotope of phosphorus, sulfur, or hydrogen, may be

incorporated into a nucleotide, which is added either after or during the synthesis of the target nucleic acid. Methods for the synthesis and purification of radiolabeled nucleic acids are well known to one of skill in the art. See, e.g., Sambrook *et al.*, 1989, in *Molecular Cloning: A Laboratory Manual*, pp 10.2-10.70, Cold Spring Harbor Laboratory Press, and the references cited therein, which are hereby incorporated by reference in their entireties.

In another embodiment, the target nucleic acid can be attached to an inorganic nanoparticle. A nanoparticle is a cluster of ions with controlled size from 0.1 to 1000 nm comprised of metals, metal oxides, or semiconductors including, but not limited to Ag_2S , ZnS , CdS , CdTe , Au , or TiO_2 . Nanoparticles have unique optical, electronic and catalytic properties relative to bulk materials which can be adjusted according to the size of the particle. Methods for the attachment of nucleic acids are well known to one of skill in the art (see, e.g., Niemeyer, 2001, *Angew. Chem. Int. Ed.* 40: 4129-4158, International Patent Publication WO/0218643, and the references cited therein, the disclosures of which are hereby incorporated by reference in their entireties).

4.3. Libraries of Small Molecules

Libraries screened using the methods of the present invention can comprise a variety of types of test compounds on solid supports. In all of the embodiments described below, all of the libraries can be synthesized on solid supports or the compounds of the library can be attached to solid supports by linkers.

In some embodiments, the test compounds are nucleic acid or peptide molecules. In a non-limiting example, peptide molecules can exist in a phage display library. In other embodiments, types of test compounds include, but are not limited to, peptide analogs including peptides comprising non-naturally occurring amino acids, e.g., D-amino acids, phosphorous analogs of amino acids, such as α -amino phosphoric acids and α -amino phosphoric acids, or amino acids having non-peptide linkages, nucleic acid analogs such as phosphorothioates and PNAs, hormones, antigens, synthetic or naturally occurring drugs, opiates, dopamine, serotonin, catecholamines, thrombin, acetylcholine, prostaglandins, organic molecules, pheromones, adenosine, sucrose, glucose, lactose and galactose. Libraries of polypeptides or proteins can also be used.

In a preferred embodiment, the combinatorial libraries are small organic molecule libraries, such as, but not limited to, benzodiazepines, isoprenoids, thiazolidinones, metathiazanones, pyrrolidines, morpholino compounds, and diazepindiones. In another embodiment, the combinatorial libraries comprise peptoids; random bio-oligomers; benzodiazepines; diversomers such as hydantoins, benzodiazepines and dipeptides;

vinylous polypeptides; nonpeptidal peptidomimetics; oligocarbamates; peptidyl phosphonates; peptide nucleic acid libraries; antibody libraries; or carbohydrate libraries. Combinatorial libraries are themselves commercially available (see, e.g., Advanced ChemTech Europe Ltd., Cambridgeshire, UK; ASINEX, Moscow Russia; BioFocus plc, Sittingbourne, UK; Bionet Research (A division of Key Organics Limited), Camelford, UK; ChemBridge Corporation, San Diego, California; ChemDiv Inc, San Diego, California; ChemRx Advanced Technologies, South San Francisco, California; ComGenex Inc., Budapest, Hungary; Evotec OAI Ltd, Abingdon, UK; IF LAB Ltd., Kiev, Ukraine; Maybridge plc, Cornwall, UK; PharmaCore, Inc., North Carolina; SIDDCO Inc, Tucson, Arizona; TimTec Inc, Newark, Delaware; Tripos Receptor Research Ltd, Bude, UK; Toslab, Ekaterinburg, Russia).

In one embodiment, the combinatorial compound library for the methods of the present invention may be synthesized. There is a great interest in synthetic methods directed toward the creation of large collections of small organic compounds, or libraries, which could be screened for pharmacological, biological or other activity (Dolle, 2001, J. Comb. Chem. 3:477-517; Hall *et al.*, 2001, *ibid.* 3:125-150; Dolle, 2000, *ibid.* 2:383-433; Dolle, 1999, *ibid.* 1:235-282). The synthetic methods applied to create vast combinatorial libraries are performed in solution or in the solid phase, *i.e.*, on a solid support. Solid-phase synthesis makes it easier to conduct multi-step reactions and to drive reactions to completion with high yields because excess reagents can be easily added and washed away after each reaction step. Solid-phase combinatorial synthesis also tends to improve isolation, purification and screening. However, the more traditional solution phase chemistry supports a wider variety of organic reactions than solid-phase chemistry. Methods and strategies for the synthesis of combinatorial libraries can be found in *A Practical Guide to Combinatorial Chemistry*, A.W. Czarnik and S.H. Dewitt, eds., American Chemical Society, 1997; *The Combinatorial Index*, B.A. Bunin, Academic Press, 1998; *Organic Synthesis on Solid Phase*, F.Z. Dörwald, Wiley-VCH, 2000; and *Solid-Phase Organic Syntheses, Vol. 1*, A.W. Czarnik, ed., Wiley Interscience, 2001.

Combinatorial compound libraries of the present invention may be synthesized using apparatuses described in US Patent No. 6,358,479 to Frisina *et al.*, U.S. Patent No. 6,190,619 to Kilcoin *et al.*, US Patent No. 6,132,686 to Gallup *et al.*, US Patent No. 6,126,904 to Zuellig *et al.*, US Patent No. 6,074,613 to Harness *et al.*, US Patent No. 6,054,100 to Stanchfield *et al.*, and US Patent No. 5,746,982 to Saneii *et al.* which are hereby incorporated by reference in their entirety. These patents describe synthesis apparatuses

capable of holding a plurality of reaction vessels for parallel synthesis of multiple discrete compounds or for combinatorial libraries of compounds.

In one embodiment, the combinatorial compound library can be synthesized in solution. The method disclosed in U.S. Patent No. 6,194,612 to Boger *et al.*, which is hereby incorporated by reference in its entirety, features compounds useful as templates for solution phase synthesis of combinatorial libraries. The template is designed to permit reaction products to be easily purified from unreacted reactants using liquid/liquid or solid/liquid extractions. The compounds produced by combinatorial synthesis using the template will preferably be small organic molecules. Some compounds in the library may mimic the effects of non-peptides or peptides. In contrast to solid phase synthesis of combinatorial compound libraries, liquid phase synthesis does not require the use of specialized protocols for monitoring the individual steps of a multistep solid phase synthesis (Egner *et al.*, 1995, J. Org. Chem. 60:2652; Anderson *et al.*, 1995, J. Org. Chem. 60:2650; Fitch *et al.*, 1994, J. Org. Chem. 59:7955; Look *et al.*, 1994, J. Org. Chem. 49:7588; Metzger *et al.*, 1993, Angew. Chem., Int. Ed. Engl. 32:894; Youngquist *et al.*, 1994, Rapid Commun. Mass Spect. 8:77; Chu *et al.*, 1995, J. Am. Chem. Soc. 117:5419; Brummel *et al.*, 1994, Science 264:399; Stevanovic *et al.*, 1993, Bioorg. Med. Chem. Lett. 3:431).

Combinatorial compound libraries useful for the methods of the present invention can be synthesized on solid supports. In one embodiment, a split synthesis method, a protocol of separating and mixing solid supports during the synthesis, is used to synthesize a library of compounds on solid supports (see Lam *et al.*, 1997, Chem. Rev. 97:41-448; Ohlmeyer *et al.*, 1993, Proc. Natl. Acad. Sci. USA 90:10922-10926 and references cited therein). Each solid support in the final library has substantially one type of test compound attached to its surface. Other methods for synthesizing combinatorial libraries on solid supports, wherein one product is attached to each support, will be known to those of skill in the art (see, e.g., Nefzi *et al.*, 1997, Chem. Rev. 97:449-472 and US Patent No. 6,087,186 to Cargill *et al.* which are hereby incorporated by reference in their entirety).

As used herein, the term "solid support" is not limited to a specific type of solid support. Rather a large number of supports are available and are known to one skilled in the art. Solid supports include silica gels, resins, derivatized plastic films, glass beads, cotton, plastic beads, polystyrene beads, doped polystyrene beads (as described by Fenniri *et al.*, 2000, J. Am. Chem. Soc. 123:8151-8152), alumina gels, and polysaccharides. A suitable solid support may be selected on the basis of desired end use and suitability for various synthetic protocols. For example, for peptide synthesis, a solid support can be a resin such as p-methylbenzhydrylamine (pMBHA) resin (Peptides International, Louisville, KY),

polystyrenes (e.g., PAM-resin obtained from Bachem Inc., Peninsula Laboratories, etc.), including chloromethylpolystyrene, hydroxymethylpolystyrene and aminomethylpolystyrene, poly (dimethylacrylamide)-grafted styrene co-divinyl-benzene (e.g., POLYHIPE resin, obtained from Aminotech, Canada), polyamide resin (obtained from Peninsula Laboratories), polystyrene resin grafted with polyethylene glycol (e.g., TENTAGEL or ARGOGEL, Bayer, Tubingen, Germany) polydimethylacrylamide resin (obtained from Milligen/Bioscience, California), or Sepharose (Pharmacia, Sweden). In another embodiment, the solid support can be a magnetic bead coated with streptavidin, such as Dynabeads Streptavidin (Dynal Biotech, Oslo, Norway).

In one embodiment, the solid phase support is suitable for *in vivo* use, i.e., it can serve as a carrier or support for administration of the test compound to a patient (e.g., TENTAGEL, Bayer, Tubingen, Germany). In a particular embodiment, the solid support is palatable and/or orally ingestible.

In some embodiments of the present invention, compounds can be attached to solid supports via linkers. Linkers can be integral and part of the solid support, or they may be nonintegral that are either synthesized on the solid support or attached thereto after synthesis. Linkers are useful not only for providing points of test compound attachment to the solid support, but also for allowing different groups of molecules to be cleaved from the solid support under different conditions, depending on the nature of the linker. For example, linkers can be, *inter alia*, electrophilically cleaved, nucleophilically cleaved, photocleavable, enzymatically cleaved, cleaved by metals, cleaved under reductive conditions or cleaved under oxidative conditions.

4.4. Library Screening

After a target nucleic acid, such as but not limited to RNA or DNA, is labeled and a test compound library is synthesized or purchased or both, the labeled target nucleic acid is used to screen the library to identify test compounds that bind to the nucleic acid. Screening comprises contacting a labeled target nucleic acid with an individual, or small group, of the components of the compound library. Preferably, the contacting occurs in an aqueous solution, and most preferably, under physiologic conditions. The aqueous solution preferably stabilizes the labeled target nucleic acid and prevents denaturation or degradation of the nucleic acid without interfering with binding of the test compounds. The aqueous solution can be similar to the solution in which a complex between the target RNA and its corresponding host cell factor is formed *in vitro*. For example, TK buffer, which is commonly used to form Tat protein-TAR RNA complexes *in vitro*, can be used in the

methods of the invention as an aqueous solution to screen a library of test compounds for TAR RNA binding compounds.

The methods of the present invention for screening a library of test compounds preferably comprise contacting a test compound with a target nucleic acid in the presence of an aqueous solution, the aqueous solution comprising a buffer and a combination of salts, preferably approximating or mimicking physiologic conditions. The aqueous solution optionally further comprises non-specific nucleic acids, such as, but not limited to, DNA; yeast tRNA; salmon sperm DNA; homoribopolymers such as, but not limited to, poly IC, polyA, polyU, and polyC; and non-specific RNA. The non-specific RNA may be an unlabeled target nucleic acid having a mutation at the binding site, which renders the unlabeled nucleic acid incapable of interacting with a test compound at that site. For example, if dye-labeled TAR RNA is used to screen a library, unlabeled TAR RNA having a mutation in the uracil 23/cytosine 24 bulge region may also be present in the aqueous solution. Without being bound by any theory, the addition of unlabeled RNA that is essentially identical to the dye-labeled target RNA except for a mutation at the binding site might minimize interactions of other regions of the dye-labeled target RNA with test compounds or with the solid support and prevent false positive results.

The solution further comprises a buffer, a combination of salts, and optionally, a detergent or a surfactant. The pH of the solution typically ranges from about 5 to about 8, preferably from about 6 to about 8, most preferably from about 6.5 to about 8. A variety of buffers may be used to achieve the desired pH. Suitable buffers include, but are not limited to, Tris, Mes, Bis-Tris, Ada, Aces, Pipes, Mopso, Bis-Tris propane, Bes, Mops, Tes, Hepes, Dipso, Mops, Tapso, Trizma, Heppso, Popso, TEA, Epps, Tricine, Gly-Gly, Bicine, and sodium-potassium phosphate. The buffering agent comprises from about 10 mM to about 100 mM, preferably from about 25 mM to about 75 mM, most preferably from about 40 mM to about 60 mM buffering agent. The pH of the aqueous solution can be optimized for different screening reactions, depending on the target RNA used and the types of test compounds in the library, and therefore, the type and amount of the buffer used in the solution can vary from screen to screen. In a preferred embodiment, the aqueous solution has a pH of about 7.4, which can be achieved using about 50 mM Tris buffer.

In addition to an appropriate buffer, the aqueous solution further comprises a combination of salts, from about 0 mM to about 100 mM KCl, from about 0 mM to about 1 M NaCl, and from about 0 mM to about 200 mM $MgCl_2$. In a preferred embodiment, the combination of salts is about 100 mM KCl, 500 mM NaCl, and 10 mM $MgCl_2$. Without being bound by any theory, Applicant has found that a combination of KCl, NaCl, and $MgCl_2$

stabilizes the target RNA such that most of the RNA is not denatured or digested over the course of the screening reaction. The optional concentration of each salt used in the aqueous solution is dependent on the particular target RNA used and can be determined using routine experimentation.

The solution optionally comprises from about 0.01% to about 0.5% (w/v) of a detergent or a surfactant. Without being bound by any theory, a small amount of detergent or surfactant in the solution might reduce non-specific binding of the target RNA to the solid support and control aggregation and increase stability of target RNA molecules. Typical detergents useful in the methods of the present invention include, but are not limited to, anionic detergents, such as salts of deoxycholic acid, 1-heptanesulfonic acid, N-laurylsarcosine, lauryl sulfate, 1-octane sulfonic acid and taurocholic acid; cationic detergents such as benzalkonium chloride, cetylpyridinium, methylbenzethonium chloride, and decamethonium bromide; zwitterionic detergents such as CHAPS, CHAPSO, alkyl betaines, alkyl amidoalkyl betaines, N-dodecyl-N,N-dimethyl-3-ammonio-1-propanesulfonate, and phosphatidylcholine; and non-ionic detergents such as n-decyl β -D-glucopyranoside, n-decyl β -D-maltopyranoside, n-dodecyl β -D-maltoside, n-octyl β -D-glucopyranoside, sorbitan esters, n-tetradecyl β -D-maltoside, octylphenoxy polyethoxyethanol (Nonidet P-40), nonylphenoxypolyethoxyethanol (NP-40), and tritons. Preferably, the detergent, if present, is a nonionic detergent. Typical surfactants useful in the methods of the present invention include, but are not limited to, ammonium lauryl sulfate, polyethylene glycols, butyl glucoside, decyl glucoside, Polysorbate 80, lauric acid, myristic acid, palmitic acid, potassium palmitate, undecanoic acid, lauryl betaine, and lauryl alcohol. More preferably, the detergent, if present, is Triton X-100 and present in an amount of about 0.1% (w/v).

Non-specific binding of a labeled target nucleic acid to test compounds can be further minimized by treating the binding reaction with one or more blocking agents. In one embodiment, the binding reactions are treated with a blocking agent, *e.g.*, bovine serum albumin ("BSA"), before contacting with the labeled target nucleic acid. In another embodiment, the binding reactions are treated sequentially with at least two different blocking agents. This blocking step is preferably performed at room temperature for from about 0.5 to about 3 hours. In a subsequent step, the reaction mixture is further treated with unlabeled RNA having a mutation at the binding site. This blocking step is preferably performed at about 4°C for from about 12 hours to about 36 hours before addition of the dye-labeled target RNA. Preferably, the solution used in the one or more blocking steps is substantially similar to the aqueous solution used to screen the library with the dye-labeled target RNA, *e.g.*, in pH and salt concentration.

Once contacted, the mixture of labeled target nucleic acid and the test compound is preferably maintained at 4°C for from about 1 day to about 5 days, preferably from about 2 days to about 3 days with constant agitation. To identify the reactions in which binding to the labeled target nucleic acid occurred, after the incubation period, bound from free compounds are determined using any of the methods disclosed in Section 4.5 *infra*.

4.5. Separation Methods for Screening Test Compounds

After the labeled target RNA is contacted with the library of test compounds immobilized on beads, the beads must then be separated from the unbound target RNA in the liquid phase. This can be accomplished by any number of physical means; *e.g.*, sedimentation, centrifugation. Thereafter, a number of methods can be used to separate the library beads that are complexed with the labeled target RNA from uncomplexed beads in order to isolate the test compound on the bead. Alternatively, mass spectroscopy and NMR spectroscopy can be used to simultaneously identify and separate beads complexed to the labeled target RNA from uncomplexed beads.

4.5.1. Flow Cytometry

In a preferred embodiment, the complexed and non-complexed target nucleic acids are separated by flow cytometry methods. Flow cytometers for sorting and examining biological cells are well known in the art; this technology can be applied to separate the labeled library beads from unlabeled beads. Known flow cytometers are described, for example, in U.S. Patent Nos. 4,347,935; 5,464,581; 5,483,469; 5,602,039; 5,643,796; and 6,211,477; the entire contents of which are incorporated by reference herein. Other known flow cytometers are the FACS Vantage™ system manufactured by Becton Dickinson and Company, and the COPAS™ system manufactured by Union Biometrica.

A flow cytometer typically includes a sample reservoir for receiving a biological sample. The biological sample contains particles (hereinafter referred to as "beads") that are to be analyzed and sorted by the flow cytometer. Beads are transported from the sample reservoir at high speed (>100beads/second) to a flow cell in a stream of liquid "sheath" fluid. High-frequency vibrations of a nozzle that directs the stream to the flow cell causes the stream to partition and form ordered droplets, with each droplet containing a single bead. Physical properties of beads can be measured as they intersect a laser beam within the cytometer flow cell. As beads move one by one through the interrogation point, they cause the laser light to scatter and fluorescent molecules on the labeled beads (*i.e.*, beads complexed with labeled target RNA) become excited.

Alternatively, if the target nucleic acid is labeled with an inorganic nanoparticle, the beads complexed with bound target nucleic acid can be distinguished not only by unique fluorescent properties but also on the basis of spectrometric properties (*e.g.* including but not limited to increased optical density due to the reduction of Ag^+ ions in the presence of gold nanoparticles (see, *e.g.*, Taton *et al.* Science 2000, 289: 1757-1760)).

An appropriate detection system consisting of photomultiplier tubes, photodiodes or other devices for measuring light are focused onto the interrogation point where the properties are measured. In so doing, information regarding particle size (light scatter) and complex formation (fluorescence intensity) is obtained. Particles with the desired physical properties are then sorted by a variety of physical means. In one embodiment, the beads are sorted by an electrostatic method. To sort beads by an electrostatic method, the droplets containing the beads with the desired physical properties are electrically charged and deflected from the trajectory of uncharged droplets as they pass through an electrostatic field formed by two deflection plates held constant at a high electrical potential difference. In another embodiment, the beads are sorted by an air-diverting method. To sort beads by an air-diverting method, the droplets containing the beads with the desired physical properties are deflected from their trajectory by a focused stream of forced air. Both of these embodiments cause the trajectory of beads with the desired physical properties to become changed, thereby sorting them from other beads. Accordingly, the beads complexed to the labeled target RNA can be collected in an appropriate collecting vessel.

Thus, in one embodiment of the present invention, the complexed and non-complexed target nucleic acids are separated by flow cytometry methods. In a preferred embodiment, the target nucleic acid is labeled with a fluorescent label and the complexed and non-complexed target nucleic acids are separated by fluorescence activated cell sorting ("FACS"). Such methods are well known to one of skill in the art.

4.5.2. Affinity Chromatography

In another embodiment of the invention, the target RNA can be labeled with biotin, an antigen, or a ligand. Library beads complexed to the target RNA can be separated from uncomplexed beads using affinity techniques designed to capture the labeled moiety on the target RNA. For example, a solid support, such as but not limited to, a column or a well in a microwell plate coated with avidin/streptavidin, an antibody to the antigen, or a receptor for the ligand can be used to capture or immobilize the labeled beads. Complexed RNA may or may not be irreversibly bound to the bead by a further transformation between the bound

RNA and an additional moiety on the surface of the bead. Such linking methods include, but are not limited to: photochemical crosslinking between RNA and bead-bound molecules such as psoralen, thymidine or uridine derivatives either present as monomers, oligomers, or as a partially complementary sequence; or chemical ligation by disulfide exchange, nitrogen mustards, bond formation between an electrophile and a nucleophile, or alkylating reagents. See, *e.g.*, International Patent Publication WO/0146461, the contents of which are hereby incorporated by reference. The unbound library beads can be removed after the binding reaction by washing the solid phase. If the RNA is irreversibly bound to the bead, test compounds can be isolated from the bead following destruction of the bound RNA by preferably, but not limited to, enzymatic or chemical (*e.g.*, alkaline hydrolysis) degradation. The library beads bound to the solid phase can then be eluted with any solution that disrupts the binding between the labeled target RNA and the solid phase. Such solutions include high salt solutions, low pH solutions, detergents, and chaotropic denaturants, and are well known to one of skill in the art. In another embodiment, the test compounds can be eluted from the solid phase by heat.

In one embodiment, the library of test compounds can be prepared on magnetic beads, such as Dynabeads Streptavidin (Dynal Biotech, Oslo, Norway). The magnetic bead library can then be mixed with the labeled target RNA under conditions that allow binding to occur. The separation of the beads from unbound target RNA in the liquid phase can be accomplished using a magnet. After removal of the magnetic field, the bead complexed to the labeled RNA may be separated from uncomplexed library beads via the label used on the target RNA; *e.g.*, biotinylated target RNA can be captured by avidin/streptavidin; target RNA labeled with antigen can be captured by the appropriate antibody; target RNA labeled with ligand can be captured using the appropriate immobilized receptor. The captured library bead can then be eluted with any solution that disrupts the binding between the labeled target RNA and the immobilized surface. Such solutions include high salt solutions, low pH solutions, detergents, and chaotropic denaturants, and are well known to one of skill in the art. Complexed RNA may or may not be irreversibly bound to the bead by a further transformation between the bound RNA and an additional moiety on the surface of the bead. Such linking methods include, but are not limited to: photochemical crosslinking between RNA and bead-bound molecules such as psoralen, thymidine or uridine derivatives either present as monomers, oligomers, or as a partially complementary sequence; or chemical ligation by disulfide exchange, nitrogen mustards, bond formation between an electrophile and a nucleophile, or alkylating reagents. See, *e.g.*, International Patent Publication WO/0146461, the contents of which are hereby incorporated by reference. If the

RNA is irreversibly bound to the bead, test compounds can be isolated from the bead following destruction of the bound RNA by enzymatic degradation including, but not limited to, ribonucleases A, U₂, CL₃, T₁, Phy M, *B. cereus* or chemical degradation including, but not limited to, piperidine-promoted backbone cleavage of abasic sites (following treatment with sodium hydroxide, hydrazine, piperidine formate, or dimethyl sulfate), or metal-assisted (*e.g.* nickel(II), cobalt(II), or iron(II)) oxidative cleavage.

In another embodiment, the preselected target RNA can be labeled with a heavy metal tag and incubated with the library beads to allow binding of the test compounds to the target RNA. The separation of the labeled beads from unlabeled beads can be accomplished using a magnetic field. After removal of the magnetic field, the test compound can be eluted with any solution that disrupts the binding between the preselected target RNA and the test compound. Such solutions include high salt solutions, low pH solutions, detergents, and chaotropic denaturants, and are well known to one of skill in the art. In another embodiment, the test compounds can be eluted from the solid phase by heat.

4.5.3. Manual Batch

In one embodiment, a manual "batch" mode is used for separating complexed beads. To explore a bead-based library within a reasonable time period, the primary screens should be operated with sufficient throughput. To do this, the target nucleic acid is labeled with a dye and then incubated with the combinatorial library. An advantage of such an assay is the fast identification of active library beads by color change. In the lower concentrations of the dye-labeled target molecule, only those library beads that bind the target molecules most tightly are detected because of higher local concentration of the dye. When washed and plated into a liquid monolayer, colored beads are easily separated from non-colored beads with the aid of a dissecting microscope. One of the problems associated with this method could be the interaction between the red dye and library substrates. Control experiments using the dye alone and dye attached to mutant RNA sequences with the libraries are performed to eliminate this possibility.

4.5.4. Suspension of Beads in Electric Fields

In another embodiment of the invention, library beads bound to the target RNA can be separated from unbound beads on the basis of the altered charge properties due to RNA binding. In a preferred embodiment of this technique, beads are separated from unbound nucleic acid and suspended, preferably but not only, in the presence of an electric field where the bound RNA causes the beads bound to the target RNA to migrate toward the

anode, or positive, end of the field.

Beads can be preferentially suspended in solution as a colloidal suspension with the aid of detergents or surfactants. Typical detergents useful in the methods of the present invention include, but are not limited to, anionic detergents, such as salts of deoxycholic acid, 1-heptanesulfonic acid, N-laurylsarcosine, lauryl sulfate, 1-octane sulfonic acid, carboxymethylcellulose, carrageenan, and taurocholic acid; cationic detergents such as benzalkonium chloride, cetylpyridinium, methylbenzethonium chloride, and decamethonium bromide; zwitterionic detergents such as CHAPS, CHAPSO, alkyl betaines, alky amidoalkyl betaines, N-dodecyl-N,N-dimethyl-3-ammonio-1-propanesulfonate, and phosphatidylcholine; and non-ionic detergents such as n-decyl α -D-glucopyranoside, n-decyl-D-maltopyranoside, n-dodecyl-D-maltoside, n-octyl-D-glucopyranoside, sorbitan esters, n-tetradecyl-D-maltoside and tritons. Preferably, the detergent, if present, is a nonionic detergent. Typical surfactants useful in the methods of the present invention include, but are not limited to, ammonium lauryl sulfate, polyethylene glycols, butyl glucoside, decyl glucoside, Polysorbate 80, lauric acid, myristic acid, palmitic acid, potassium palmitate, undecanoic acid, lauryl betaine, and lauryl alcohol.

Complexed RNA may or may not be irreversibly bound to the bead by a further transformation between the bound RNA and an additional moiety on the surface of the bead. Such linking methods include, but are not limited to: photochemical crosslinking between RNA and bead-bound molecules such as psoralen, thymidine or uridine derivatives either present as monomers, oligomers, or as a partially complementary sequence; or chemical ligation by disulfide exchange, nitrogen mustards, bond formation between an electrophile and a nucleophile, or alkylating reagents.

If the RNA is irreversibly bound to the bead, test compounds can be isolated from the bead following destruction of the bound RNA by enzymatic degradation including, but not limited to, ribonucleases A, U₂, C₁, T₁, Phy M, *B. cereus* or chemical degradation including, but not limited to, piperidine-promoted backbone cleavage of abasic sites (following treatment with sodium hydroxide, hydrazine, piperidine formate, or dimethyl sulfate), or metal-assisted (e.g. nickel(II), cobalt(II), or iron(II)) oxidative cleavage.

4.5.5. Microwave

In another embodiment, the complexed beads are separated from uncomplexed beads by microwave. For example, as described in U.S. Patent Nos. 6,340,568; 6,338,968; and 6,287,874 to Hefti, the disclosures of which are hereby incorporated by reference, a system which is sensitive to the unique dielectric properties of

molecules and binding complexes, such as hybridization complexes formed between a nucleic acid probe and a nucleic acid target, molecular binding events, and protein/ligand complexes, can be used to analyze nucleic acids. In this system, the different hybridization complexes can be directly distinguished without the use of labels. The method involves contacting a nucleic acid probe that is electromagnetically coupled to a portion of a signal path with a sample containing a target nucleic acid. The portion of the signal path to which the nucleic acid probe is coupled typically is a continuous transmission line. A response signal is detected for a hybridization complex formed between the nucleic acid probe and the nucleic acid target. Detection may involve propagating a test signal along the signal path and then detecting a response signal formed through modulation of the test signal by the hybridization complex.

4.6. Methods for Identifying Test Compounds

If the library is a peptide or nucleic acid library, the sequence of the test compound on the isolated bead can be determined by direct sequencing of the peptide or nucleic acid. Such methods are well known to one of skill in the art.

4.6.1. Mass Spectrometry

Mass spectrometry (*e.g.*, electrospray ionization ("ESI") and matrix-assisted laser desorption-ionization ("MALDI"), Fourier-transform ion cyclotron resonance ("FT-ICR")) can be used both for high-throughput screening of test compounds that bind to a target RNA and elucidating the structure of the test compound on the isolated bead.

MALDI uses a pulsed laser for desorption of the ions and a time-of-flight analyzer, and has been used for the detection of noncovalent tRNA:amino-acyl-tRNA synthetase complexes (Gruic-Sovulj *et al.*, 1997, J. Biol. Chem. 272:32084-32091). However, covalent cross-linking between the target nucleic acid and the test compound is required for detection, since a non-covalently bound complex may dissociate during the MALDI process.

ESI mass spectrometry ("ESI-MS") has been of greater utility for studying non-covalent molecular interactions because, unlike the MALDI process, ESI-MS generates molecular ions with little to no fragmentation (Xavier *et al.*, 2000, Trends Biotechnol. 18(8):349-356). ESI-MS has been used to study the complexes formed by HIV Tat peptide and protein with the TAR RNA (Sannes-Lowery *et al.*, 1997, Anal. Chem. 69:5130-5135).

Fourier-transform ion cyclotron resonance ("FT-ICR") mass spectrometry provides high-resolution spectra, isotope-resolved precursor ion selection, and accurate mass

assignments (Xavier *et al.*, 2000, Trends Biotechnol. 18(8):349-356). FT-ICR has been used to study the interaction of aminoglycoside antibiotics with cognate and non-cognate RNAs (Hofstadler *et al.*, 1999, Anal. Chem. 71:3436-3440; Griffey *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96:10129-10133). As true for all of the mass spectrometry methods discussed herein, FT-ICR does not require labeling of the target RNA or a test compound.

An advantage of mass spectroscopy is not only the elucidation of the structure of the test compound, but also the determination of the structure of the test compound bound to the preselected target RNA. Such information can enable the discovery of a consensus structure of a test compound that specifically binds to a preselected target RNA.

In a preferred embodiment, the structure of the test compound is determined by time of flight mass spectroscopy ("TOF-MS"). In time of flight methods of mass spectrometry, charged (ionized) molecules are produced in a vacuum and accelerated by an electric field into a time of flight tube or drift tube. The velocity to which the molecules may be accelerated is proportional to the accelerating potential, proportional to the charge of the molecule, and inversely proportional to the square of the mass of the molecule. The charged molecules travel, *i.e.*, "drift" down the TOF tube to a detector. The time taken for the molecules to travel down the tube may be interpreted as a measure of their molecular weight. Time-of-flight mass spectrometers have been developed for all of the major ionization techniques such as, but limited to, electron impact ("EI"), infrared laser desorption ("IRLD"), plasma desorption ("PD"), fast atom bombardment ("FAB"), secondary ion mass spectrometry ("SIMS"), matrix-assisted laser desorption/ionization ("MALDI"), and electrospray ionization ("ESI").

4.6.2. NMR Spectroscopy

NMR spectroscopy can be used for elucidating the structure of the test compound on the isolated bead. NMR spectroscopy is a technique for identifying binding sites in target nucleic acids by qualitatively determining changes in chemical shift, specifically from distances measured using relaxation effects. Examples of NMR that can be used for the invention include, but are not limited to, one-dimensional NMR, two-dimensional NMR, correlation spectroscopy ("COSY"), and nuclear Overhauser effect ("NOE") spectroscopy. Such methods of structure determination of test compounds are well known to one of skill in the art.

Similar to mass spectroscopy, an advantage of NMR is the not only the elucidation of the structure of the test compound, but also the determination of the structure of the test compound bound to the preselected target RNA. Such information can enable the

discovery of a consensus structure of a test compound that specifically binds to a preselected target RNA.

4.6.3. Edman Degradation

5 In an embodiment wherein the library is a peptide library or a derivative thereof, Edman degradation can be used to determine the structure of the test compound. In one embodiment, a modified Edman degradation process is used to obtain compositional tags for proteins, which is described in U.S. Patent No. 6,277,644 to Farnsworth *et al.*, which is hereby incorporated by reference in its entirety. The Edman degradation chemistry is
10 separated from amino acid analysis, circumventing the serial requirement of the conventional Edman process. Multiple cycles of coupling and cleavage are performed prior to extraction and compositional analysis of amino acids. The amino acid composition information is then used to search a database of known protein or DNA sequences to identify the sample protein.
15 An apparatus for performing this method comprises a sample holder for holding the sample, a coupling agent supplier for supplying at least one coupling agent, a cleavage agent supplier for supplying a cleavage agent, a controller for directing the sequential supply of the coupling agents, cleavage agents, and other reagents necessary for performing the modified Edman degradation reactions, and an analyzer for analyzing amino acids.

20 In another embodiment, the method can be automated as described in U.S. Patent No. 5,565,171 to Dovichi *et al.*, which is hereby incorporated by reference in its entirety. The apparatus includes a continuous capillary connected between two valves that control fluid flow in the capillary. One part of the capillary forms a reaction chamber where the sample may be immobilized for subsequent reaction with reagents supplied through the
25 valves. Another part of the capillary passes through or terminates in the detector portion of an analyzer such as an electrophoresis apparatus, liquid chromatographic apparatus or mass spectrometer. The apparatus may form a peptide or protein sequencer for carrying out the Edman degradation reaction and analyzing the reaction product produced by the reaction. The protein or peptide sequencer includes a reaction chamber for carrying out coupling and
30 cleavage on a peptide or protein to produce derivatized amino acid residue, a conversion chamber for carrying out conversion and producing a converted amino acid residue and an analyzer for identifying the converted amino acid residue. The reaction chamber may be contained within one arm of a capillary and the conversion chamber is located in another arm of the capillary. An electrophoresis length of capillary is directly capillary coupled to the
35 conversion chamber to allow electrophoresis separation of the converted amino acid residue

as it leaves the conversion chamber. Identification of the converted amino acid residue takes place at one end of the electrophoresis length of the capillary.

4.6.4. Vibrational Spectroscopy

Vibrational spectroscopy (*e.g.* infrared (IR) spectroscopy or Raman spectroscopy) can be used for elucidating the structure of the test compound on the isolated bead.

Infrared spectroscopy measures the frequencies of infrared light (wavelengths from 100 to 10,000 nm) absorbed by the test compound as a result of excitation of vibrational modes according to quantum mechanical selection rules which require that absorption of light cause a change in the electric dipole moment of the molecule. The infrared spectrum of any molecule is a unique pattern of absorption wavelengths of varying intensity that can be considered as a molecular fingerprint to identify any compound.

Infrared spectra can be measured in a scanning mode by measuring the absorption of individual frequencies of light, produced by a grating which separates frequencies from a mixed-frequency infrared light source, by the test compound relative to a standard intensity (double-beam instrument) or pre-measured ('blank') intensity (single-beam instrument). In a preferred embodiment, infrared spectra are measured in a pulsed mode (FT-IR) where a mixed beam, produced by an interferometer, of all infrared light frequencies is passed through or reflected off the test compound. The resulting interferogram, which may or may not be added with the resulting interferograms from subsequent pulses to increase the signal strength while averaging random noise in the electronic signal, is mathematically transformed into a spectrum using Fourier Transform or Fast Fourier Transform algorithms.

Raman spectroscopy measures the difference in frequency due to absorption of infrared frequencies of scattered visible or ultraviolet light relative to the incident beam. The incident monochromatic light beam, usually a single laser frequency, is not truly absorbed by the test compound but interacts with the electric field transiently. Most of the light scattered off the sample will be unchanged (Rayleigh scattering) but a portion of the scatter light will have frequencies that are the sum or difference of the incident and molecular vibrational frequencies. The selection rules for Raman (inelastic) scattering require a change in polarizability of the molecule. While some vibrational transitions are observable in both infrared and Raman spectrometry, many are observable only with one or the other technique. The Raman spectrum of any molecule is a unique pattern of absorption wavelengths of varying intensity that can be considered as a molecular fingerprint to identify any compound.

Raman spectra are measured by submitting monochromatic light to the sample, either passed through or preferably reflected off, filtering the Rayleigh scattered light, and detecting the frequency of the Raman scattered light. An improved Raman spectrometer is described in US Patent No. 5,786,893 to Fink *et al.*, which is hereby
5 incorporated by reference.

Vibrational microscopy can be measured in a spatially resolved fashion to address single beads by integration of a visible microscope and spectrometer. A microscopic infrared spectrometer is described in U.S. Patent No. 5,581,085 to Reffner *et al.*, which is
10 hereby incorporated by reference in its entirety. An instrument that simultaneously performs a microscopic infrared and microscopic Raman analysis on a sample is described in U.S. Patent No. 5,841,139 to Sostek *et al.*, which is hereby incorporated by reference in its entirety.

In one embodiment of the method, test compounds are synthesized on polystyrene beads doped with chemically modified styrene monomers such that each
15 resulting bead has a characteristic pattern of absorption lines in the vibrational (IR or Raman) spectrum, by methods including but not limited to those described by Fenniri *et al.*, 2000, J. Am. Chem. Soc. 123:8151-8152. Using methods of split-pool synthesis familiar to one of skill in the art, the library of compounds is prepared so that the spectroscopic pattern of the
20 bead identifies one of the components of the test compound on the bead. Beads that have been separated according to their ability to bind target RNA can be identified by their vibrational spectrum. In one embodiment of the method, appropriate sorting and binning of the beads during synthesis then allows identification of one or more further components of the test compound on any one bead. In another embodiment of the method, partial
25 identification of the compound on a bead is possible through use of the spectroscopic pattern of the bead with or without the aid of further sorting during synthesis, followed by partial resynthesis of the possible compounds aided by doped beads and appropriate sorting during synthesis.

In another embodiment, the IR or Raman spectra of test compounds are
30 examined while the compound is still on a bead, preferably, or after cleavage from bead, using methods including but not limited to photochemical, acid, or heat treatment. The test compound can be identified by comparison of the IR or Raman spectral pattern to spectra previously acquired for each test compound in the combinatorial library.

4.7. Secondary Biological Screens

The test compounds identified in the binding assay (for convenience referred to herein as a "lead" compound) can be tested for biological activity using host cells containing or engineered to contain the target RNA element coupled to a functional readout system. For example, the lead compound can be tested in a host cell engineered to contain the target RNA element controlling the expression of a reporter gene. In this example, the lead compounds are assayed in the presence or absence of the target RNA. Alternatively, a phenotypic or physiological readout can be used to assess activity of the target RNA in the presence and absence of the lead compound.

In one embodiment, the lead compound can be tested in a host cell engineered to contain the target RNA element controlling the expression of a reporter gene, such as, but not limited to, β -galactosidase, green fluorescent protein, red fluorescent protein, luciferase, chloramphenicol acetyltransferase, alkaline phosphatase, and β -lactamase. In a preferred embodiment, a cDNA encoding the target element is fused upstream to a reporter gene wherein translation of the reporter gene is repressed upon binding of the lead compound to the target RNA. In other words, the steric hindrance caused by the binding of the lead compound to the target RNA repressed the translation of the reporter gene. This method, termed the translational repression assay procedure ("TRAP") has been demonstrated in *E. coli* and *S. cerevisiae* (Jain & Belasco, 1996, Cell 87(1):115-25; Huang & Schreiber, 1997, Proc. Natl. Acad. Sci. USA 94:13396-13401).

In another embodiment, a phenotypic or physiological readout can be used to assess activity of the target RNA in the presence and absence of the lead compound. For example, the target RNA may be overexpressed in a cell in which the target RNA is endogenously expressed. Where the target RNA controls expression of a gene product involved in cell growth or viability, the *in vivo* effect of the lead compound can be assayed by measuring the cell growth or viability of the target cell. Alternatively, a reporter gene can also be fused downstream of the target RNA sequence and the effect of the lead compound on reporter gene expression can be assayed.

Alternatively, the lead compounds identified in the binding assay can be tested for biological activity using animal models for a disease, condition, or syndrome of interest. These include animals engineered to contain the target RNA element coupled to a functional readout system, such as a transgenic mouse. Animal model systems can also be used to demonstrate safety and efficacy.

Compounds displaying the desired biological activity can be considered to be lead compounds, and will be used in the design of congeners or analogs possessing useful

pharmacological activity and physiological profiles. Following the identification of a lead compound, molecular modeling techniques can be employed, which have proven to be useful in conjunction with synthetic efforts, to design variants of the lead that can be more effective.

- 5 These applications may include, but are not limited to, Pharmacophore Modeling (*cf.* Lamothe, *et al.* 1997, *J. Med. Chem.* 40: 3542; Mottola *et al.* 1996, *J. Med. Chem.* 39: 285; Beusen *et al.* 1995, *Biopolymers* 36: 181; P. Fossa *et al.* 1998, *Comput. Aided Mol. Des.* 12: 361), QSAR development (*cf.* Siddiqui *et al.* 1999, *J. Med. Chem.* 42: 4122; Barreca *et al.* 1999 *Bioorg. Med. Chem.* 7: 2283; Kroemer *et al.* 1995, *J. Med. Chem.* 38: 4917; Schaal *et al.* 10 2001, *J. Med. Chem.* 44: 155; Buolamwini & Assefa 2002, *J. Mol. Chem.* 45: 84), Virtual docking and screening/scoring (*cf.* Anzini *et al.* 2001, *J. Med. Chem.* 44: 1134; Faaland *et al.* 2000, *Biochem. Cell. Biol.* 78: 415; Silvestri *et al.* 2000, *Bioorg. Med. Chem.* 8: 2305; J. Lee *et al.* 2001, *Bioorg. Med. Chem.* 9: 19), and Structure Prediction using RNA structural programs including, but not limited to mFold (as described by Zuker *et al.* Algorithms and
15 Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide in RNA Biochemistry and Biotechnology pp. 11-43, J. Barciszewski & B.F.C. Clark, eds. (NATO ASI Series, Kluwer Academic Publishers, 1999) and Mathews *et al.* 1999 *J. Mol. Biol.* 288: 911-940); RNAmotif (Macke *et al.* 2001, *Nucleic Acids Res.* 29: 4724-4735; and the Vienna RNA package (Hofacker *et al.* 1994, *Monatsh. Chem.* 125: 167-188).

- 20 Further examples of the application of such techniques can be found in several review articles, such as Rotivinen *et al.*, 1988, *Acta Pharmaceutica Fennica* 97:159-166; Ripka, 1998, *New Scientist* 54-57; McKinaly & Rossmann, 1989, *Annu. Rev. Pharmacol. Toxicol.* 29:111-122; Perry & Davies, QSAR: Quantitative Structure-Activity Relationships in Drug Design pp. 189-193 (Alan R. Liss, Inc. 1989); Lewis & Dean, 1989, *Proc. R. Soc.*
25 *Lond.* 236:125-140 and 141-162; Askew *et al.*, 1989, *J. Am. Chem. Soc.* 111:1082-1090. Molecular modeling tools employed may include those from Tripos, Inc., St. Louis, Missouri (*e.g.*, Sybyl/UNITY, CONCORD, DiverseSolutions), Accelrys, San Diego, California (*e.g.*, Catalyst, Wisconsin Package {BLAST, etc.}), Schrodinger, Portland, Oregon (*e.g.*, QikProp, QikFit, Jaguar) or other such vendors as BioDesign, Inc. (Pasadena, California), Allelix, Inc.
30 (Mississauga, Ontario, Canada), and Hypercube, Inc. (Cambridge, Ontario, Canada), and may include privately designed and/or "academic" software (*e.g.* RNAmotif, mFOLD). These application suites and programs include tools for the atomistic construction and analysis of structural models for drug-like molecules, proteins, and DNA or RNA and their potential interactions. They also provide for the calculation of important physical properties, such as
35 solubility estimates, permeability metrics, and empirical measures of molecular "druggability" (*e.g.*, Lipinski "Rule of 5" as described by Lipinski *et al.* 1997, *Adv. Drug*

Delivery Rev. 23: 3-25). Most importantly, they provide appropriate metrics and statistical modeling power (such as the patented CoMFA technology in Sybyl as described in US Patents 6,240,374 and 6,185,506) to develop Quantitative Structural Activity Relationships (QSARs) which are used to guide the synthesis of more efficacious clinical development candidates while improving desirable physical properties, as determined by results from the aforementioned secondary screening protocols.

4.8. Use of Identified Compounds That Bind RNA to Treat/Prevent Disease

Biologically active compounds identified using the methods of the invention or a pharmaceutically acceptable salt thereof can be administered to a patient, preferably a mammal, more preferably a human, suffering from a disease whose progression is associated with a target RNA:host cell factor interaction *in vivo*. In certain embodiments, such compounds or a pharmaceutically acceptable salt thereof is administered to a patient, preferably a mammal, more preferably a human, as a preventative measure against a disease associated with an RNA:host cell factor interaction *in vivo*.

In one embodiment, "treatment" or "treating" refers to an amelioration of a disease, or at least one discernible symptom thereof. In another embodiment, "treatment" or "treating" refers to an amelioration of at least one measurable physical parameter, not necessarily discernible by the patient. In yet another embodiment, "treatment" or "treating" refers to inhibiting the progression of a disease, either physically, *e.g.*, stabilization of a discernible symptom, physiologically, *e.g.*, stabilization of a physical parameter, or both. In yet another embodiment, "treatment" or "treating" refers to delaying the onset of a disease.

In certain embodiments, the compound or a pharmaceutically acceptable salt thereof is administered to a patient, preferably a mammal, more preferably a human, as a preventative measure against a disease associated with an RNA:host cell factor interaction *in vivo*. As used herein, "prevention" or "preventing" refers to a reduction of the risk of acquiring a disease. In one embodiment, the compound or a pharmaceutically acceptable salt thereof is administered as a preventative measure to a patient. According to this embodiment, the patient can have a genetic predisposition to a disease, such as a family history of the disease, or a non-genetic predisposition to the disease. Accordingly, the compound and pharmaceutically acceptable salts thereof can be used for the treatment of one manifestation of a disease and prevention of another.

When administered to a patient, the compound or a pharmaceutically acceptable salt thereof is preferably administered as component of a composition that optionally comprises a pharmaceutically acceptable vehicle. The composition can be

administered orally, or by any other convenient route, for example, by infusion or bolus injection, by absorption through epithelial or mucocutaneous linings (*e.g.*, oral mucosa, rectal, and intestinal mucosa, *etc.*) and may be administered together with another biologically active agent. Administration can be systemic or local. Various delivery systems are known, *e.g.*, encapsulation in liposomes, microparticles, microcapsules, capsules, *etc.*, and can be used to administer the compound and pharmaceutically acceptable salts thereof.

Methods of administration include but are not limited to intradermal, intramuscular, intraperitoneal, intravenous, subcutaneous, intranasal, epidural, oral, sublingual, intranasal, intracerebral, intravaginal, transdermal, rectally, by inhalation, or topically, particularly to the ears, nose, eyes, or skin. The mode of administration is left to the discretion of the practitioner. In most instances, administration will result in the release of the compound or a pharmaceutically acceptable salt thereof into the bloodstream.

In specific embodiments, it may be desirable to administer the compound or a pharmaceutically acceptable salt thereof locally. This may be achieved, for example, and not by way of limitation, by local infusion during surgery, topical application, *e.g.*, in conjunction with a wound dressing after surgery, by injection, by means of a catheter, by means of a suppository, or by means of an implant, said implant being of a porous, non-porous, or gelatinous material, including membranes, such as sialastic membranes, or fibers.

In certain embodiments, it may be desirable to introduce the compound or a pharmaceutically acceptable salt thereof into the central nervous system by any suitable route, including intraventricular, intrathecal and epidural injection. Intraventricular injection may be facilitated by an intraventricular catheter, for example, attached to a reservoir, such as an Ommaya reservoir.

Pulmonary administration can also be employed, *e.g.*, by use of an inhaler or nebulizer, and formulation with an aerosolizing agent, or via perfusion in a fluorocarbon or synthetic pulmonary surfactant. In certain embodiments, the compound and pharmaceutically acceptable salts thereof can be formulated as a suppository, with traditional binders and vehicles such as triglycerides.

In another embodiment, the compound and pharmaceutically acceptable salts thereof can be delivered in a vesicle, in particular a liposome (see Langer, 1990, *Science* 249:1527-1533; Treat *et al.*, in *Liposomes in the Therapy of Infectious Disease and Cancer*, Lopez-Berestein and Fidler (eds.), Liss, New York, pp. 353-365 (1989); Lopez-Berestein, *ibid.*, pp. 317-327; see generally *ibid.*).

In yet another embodiment, the compound and pharmaceutically acceptable salts thereof can be delivered in a controlled release system (see, *e.g.*, Goodson, in *Medical*

Applications of Controlled Release, *supra*, vol. 2, pp. 115-138 (1984)). Other controlled-release systems discussed in the review by Langer, 1990, *Science* 249:1527-1533) may be used. In one embodiment, a pump may be used (see Langer, *supra*; Sefton, 1987, *CRC Crit. Ref. Biomed. Eng.* 14:201; Buchwald *et al.*, 1980, *Surgery* 88:507; Saudek *et al.*, 1989, *N. Engl. J. Med.* 321:574). In another embodiment, polymeric materials can be used (see Medical Applications of Controlled Release, Langer and Wise (eds.), CRC Pres., Boca Raton, Florida (1974); Controlled Drug Bioavailability, Drug Product Design and Performance, Smolen and Ball (eds.), Wiley, New York (1984); Ranger and Peppas, 1983, *J. Macromol. Sci. Rev. Macromol. Chem.* 23:61; see also Levy *et al.*, 1985, *Science* 228:190; During *et al.*, 1989, *Ann. Neurol.* 25:351; Howard *et al.*, 1989, *J. Neurosurg.* 71:105). In yet another embodiment, a controlled-release system can be placed in proximity of a target RNA of the compound or a pharmaceutically acceptable salt thereof, thus requiring only a fraction of the systemic dose.

Compositions comprising the compound or a pharmaceutically acceptable salt thereof ("compound compositions") can additionally comprise a suitable amount of a pharmaceutically acceptable vehicle so as to provide the form for proper administration to the patient.

In a specific embodiment, the term "pharmaceutically acceptable" means approved by a regulatory agency of the Federal or a state government or listed in the U.S. Pharmacopeia or other generally recognized pharmacopeia for use in animals, mammals, and more particularly in humans. The term "vehicle" refers to a diluent, adjuvant, excipient, or carrier with which a compound of the invention is administered. Such pharmaceutical vehicles can be liquids, such as water and oils, including those of petroleum, animal, vegetable or synthetic origin, such as peanut oil, soybean oil, mineral oil, sesame oil and the like. The pharmaceutical vehicles can be saline, gum acacia, gelatin, starch paste, talc, keratin, colloidal silica, urea, and the like. In addition, auxiliary, stabilizing, thickening, lubricating and coloring agents may be used. When administered to a patient, the pharmaceutically acceptable vehicles are preferably sterile. Water is a preferred vehicle when the compound of the invention is administered intravenously. Saline solutions and aqueous dextrose and glycerol solutions can also be employed as liquid vehicles, particularly for injectable solutions. Suitable pharmaceutical vehicles also include excipients such as starch, glucose, lactose, sucrose, gelatin, malt, rice, flour, chalk, silica gel, sodium stearate, glycerol monostearate, talc, sodium chloride, dried skim milk, glycerol, propylene, glycol, water, ethanol and the like. Compound compositions, if desired, can also contain minor amounts of wetting or emulsifying agents, or pH buffering agents.

Compound compositions can take the form of solutions, suspensions, emulsion, tablets, pills, pellets, capsules, capsules containing liquids, powders, sustained-release formulations, suppositories, emulsions, aerosols, sprays, suspensions, or any other form suitable for use. In one embodiment, the pharmaceutically acceptable vehicle is a capsule (see *e.g.*, U.S. Patent No. 5,698,155). Other examples of suitable pharmaceutical vehicles are described in Remington's Pharmaceutical Sciences, Alfonso R. Gennaro, ed., Mack Publishing Co. Easton, PA, 19th ed., 1995, pp. 1447 to 1676, incorporated herein by reference.

In a preferred embodiment, the compound or a pharmaceutically acceptable salt thereof is formulated in accordance with routine procedures as a pharmaceutical composition adapted for oral administration to human beings. Compositions for oral delivery may be in the form of tablets, lozenges, aqueous or oily suspensions, granules, powders, emulsions, capsules, syrups, or elixirs, for example. Orally administered compositions may contain one or more agents, for example, sweetening agents such as fructose, aspartame or saccharin; flavoring agents such as peppermint, oil of wintergreen, or cherry; coloring agents; and preserving agents, to provide a pharmaceutically palatable preparation. Moreover, where in tablet or pill form, the compositions can be coated to delay disintegration and absorption in the gastrointestinal tract thereby providing a sustained action over an extended period of time. Selectively permeable membranes surrounding an osmotically active driving compound are also suitable for orally administered compositions. In these later platforms, fluid from the environment surrounding the capsule is imbibed by the driving compound, which swells to displace the agent or agent composition through an aperture. These delivery platforms can provide an essentially zero order delivery profile as opposed to the spiked profiles of immediate release formulations. A time delay material such as glycerol monostearate or glycerol stearate may also be used. Oral compositions can include standard vehicles such as mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like. Such vehicles are preferably of pharmaceutical grade. Typically, compositions for intravenous administration comprise sterile isotonic aqueous buffer. Where necessary, the compositions may also include a solubilizing agent.

In another embodiment, the compound or a pharmaceutically acceptable salt thereof can be formulated for intravenous administration. Compositions for intravenous administration may optionally include a local anesthetic such as lignocaine to lessen pain at the site of the injection. Generally, the ingredients are supplied either separately or mixed together in unit dosage form, for example, as a dry lyophilized powder or water-free

concentrate in a hermetically sealed container such as an ampoule or sachette indicating the quantity of active agent. Where the compound or a pharmaceutically acceptable salt thereof is to be administered by infusion, it can be dispensed, for example, with an infusion bottle containing sterile pharmaceutical grade water or saline. Where the compound or a pharmaceutically acceptable salt thereof is administered by injection, an ampoule of sterile water for injection or saline can be provided so that the ingredients may be mixed prior to administration.

The amount of a compound or a pharmaceutically acceptable salt thereof that will be effective in the treatment of a particular disease will depend on the nature of the disease, and can be determined by standard clinical techniques. In addition, *in vitro* or *in vivo* assays may optionally be employed to help identify optimal dosage ranges. The precise dose to be employed will also depend on the route of administration, and the seriousness of the disease, and should be decided according to the judgment of the practitioner and each patient's circumstances. However, suitable dosage ranges for oral administration are generally about 0.001 milligram to about 200 milligrams of a compound or a pharmaceutically acceptable salt thereof per kilogram body weight per day. In specific preferred embodiments of the invention, the oral dose is about 0.01 milligram to about 100 milligrams per kilogram body weight per day, more preferably about 0.1 milligram to about 75 milligrams per kilogram body weight per day, more preferably about 0.5 milligram to 5 milligrams per kilogram body weight per day. The dosage amounts described herein refer to total amounts administered; that is, if more than one compound is administered, or if a compound is administered with a therapeutic agent, then the preferred dosages correspond to the total amount administered. Oral compositions preferably contain about 10% to about 95% active ingredient by weight.

Suitable dosage ranges for intravenous (i.v.) administration are about 0.01 milligram to about 100 milligrams per kilogram body weight per day, about 0.1 milligram to about 35 milligrams per kilogram body weight per day, and about 1 milligram to about 10 milligrams per kilogram body weight per day. Suitable dosage ranges for intranasal administration are generally about 0.01 pg/kg body weight per day to about 1 mg/kg body weight per day. Suppositories generally contain about 0.01 milligram to about 50 milligrams of a compound of the invention per kilogram body weight per day and comprise active ingredient in the range of about 0.5% to about 10% by weight.

Recommended dosages for intradermal, intramuscular, intraperitoneal, subcutaneous, epidural, sublingual, intracerebral, intravaginal, transdermal administration or administration by inhalation are in the range of about 0.001 milligram to about 200

milligrams per kilogram of body weight per day. Suitable doses for topical administration are in the range of about 0.001 milligram to about 1 milligram, depending on the area of administration. Effective doses may be extrapolated from dose-response curves derived from *in vitro* or animal model test systems. Such animal models and systems are well known in the art.

The compound and pharmaceutically acceptable salts thereof are preferably assayed *in vitro* and *in vivo*, for the desired therapeutic or prophylactic activity, prior to use in humans. For example, *in vitro* assays can be used to determine whether it is preferable to administer the compound, a pharmaceutically acceptable salt thereof, and/or another therapeutic agent. Animal model systems can be used to demonstrate safety and efficacy.

A variety of compounds can be used for treating or preventing diseases in mammals. Types of compounds include, but are not limited to, peptides, peptide analogs including peptides comprising non-natural amino acids, *e.g.*, D-amino acids, phosphorous analogs of amino acids, such as α -amino phosphonic acids and α -amino phosphinic acids, or amino acids having non-peptide linkages, nucleic acids, nucleic acid analogs such as phosphorothioates or peptide nucleic acids ("PNAs"), hormones, antigens, synthetic or naturally occurring drugs, opiates, dopamine, serotonin, catecholamines, thrombin, acetylcholine, prostaglandins, organic molecules, pheromones, adenosine, sucrose, glucose, lactose and galactose.

5. EXAMPLE: THERAPEUTIC TARGETS

The therapeutic targets presented herein are by way of example, and the present invention is not to be limited by the targets described herein. The therapeutic targets presented herein as DNA sequences are understood by one of skill in the art that the sequences can be converted to RNA sequences.

5.1. Tumor Necrosis Factor Alpha ("TNF- α ")

GenBank Accession # X01394:

```

1 gcagaggacc agctaagagg gagagaagca actacagacc cccctgaa aacacctca
61 gacgccacat cccctgacaa gctgccagcg aggtctctt ctctcacat actgaccac
121 ggctccacc ttcttccctt ggaaggaca ccatgagcac tgaagcatg atccgggacg
181 tggagctggc cgaggagcgc ctcccaaga agacaggggg gccccagggc tcaggcggt
241 gctgttctt cagctcttc tcttctga tctggcagg cgccaccacg ctctctgcc
301 tgctgcactt tggagtgatc ggcgccaga gggaagagt cccagggac ctctcttaa
361 tcagccctct ggcgcaggca gtcagatcat ctctcgaac cccgagtgac aagcctgtag

```

421 cccatgttgt agcaaacct caagctgagg ggcagctcca gtgctgaac cgcggggcca
 481 atgcctcct ggccaatggc gtggagctga gagataacca gctgtgtgtg ccacagagg
 541 gcctgtacct catctactoc caggctctct tcaagggcc aggcgcctcc tccacctatg
 5 601 tgcctctcac ccacaccatc agccgcatcg ccgtctcta ccagaccaag gtcaacctcc
 661 tctctccat caagagcccc tgcagagggg agaccccaga gggggctgag gccaaagcct
 721 ggtatgagcc catctatctg ggaggggctct tccagctgga gaagggtgac cgactcagcg
 781 ctgagatcaa tcggcccgac tatctcgact ttgccgagtc tggcgagctc tactttggga
 841 tcattgccct gtgaggagga cgaacatcca acctcccaa acgcctcccc tgcaccaatc
 901 cctttattac cccctccttc agacacccct aacctctctt ggctcaaaaa gagaattggg
 961 ggcttagggg cggaaoccaa gcttagaact tlaagcaaca agaccaccac tlcgaaacct
 1021 gggattcagg aatgtgtggc ctgcacagtg aattgctgac aaccactaag aattcaacct
 1081 ggggcctoca gaactcactg gggcctacag ctttagcccc tgacatcgg aatctggaga
 1141 ccaggaggcc ttgtgtctg gccagaatgc tgcaggactt gagaagacct cactagaaa
 1201 tgacacaag tggaccttag gcctctctct ctccagatgt ttccagactt ccttgagaca
 1261 cggagccag cctcccat ggagccagct cctctattt atgttgac ttgtgattat
 1321 ttattattta ttattattt attatttac agatgaatgt attatttgg gagaccgggg
 1381 tctctgggg gacccaatgt aggagctgcc ttggctcaga catgtttcc gtgaaaaggg
 1441 agctgaacaa taggctgttc ccatgtagcc ccttggcctc tgtgcctct ttgattatg
 1501 tttttaaaa tatttatctg attaatgtt ctaacaatg ctgatttgt gaccaactgt
 1561 cactcattgc tgagcctctg ctcccagggg gagttgtgc tgaatcgcc ctactattca
 1621 gtggcgagaa ataaagtttg ctt (SEQ ID NO: 6)

General Target Regions:

- 25 (1) 5' Untranslated Region - nts 1 - 152
 (2) 3' Untranslated Region - nts 852 - 1643

Initial Specific Target Motif:

- Group I AU-Rich Element (ARE) Cluster in 3' untranslated region
 30 5' AUUUUUUUUUUUUUUUUUUA 3' (SEQ ID NO: 1)

5.2. Granulocyte-macrophage Colony Stimulating Factor ("GM-CSF")

GenBank Accession # NM_000758:

1 gctggaggat gtggctgcag agcctgctgc tctgggcac tgggcctgc agcatctcg
 61 caccgcccgc ctgcgccagc ccagcacgc agccctggga gcatgtgaat gccatccagg
 121 aggcccggcg tctcctgaac ctgagtagag acactgctgc tgagatgaat gaaacagtag

181 aagtcattctc agaaatgttt gacctccagg agccgacctg cctacagacc cgccctggagc
 241 tgtacaagca gggcctgcgg ggcagcctca ccaagctcaa gggcccttg accatgatgg
 301 ccagccacta caagcagcac tgcctccaa ccccgaaac ttctgtgca acccagacta
 5 361 tcaccttga aagtttcaa gagaacctga aggactttct gcttgctc ccccttgact
 421 gctgggagcc agtccaggag tgagaccggc cagatgaggc tggccaagcc ggggagctgc
 481 tctctcatga aacaagagct agaaactcag gatgtgctc ttgaggggac caagggtg
 541 gccacagcca tgggtggagt ggcctggacc tgcctgggc cacactgacc ctgatacagg
 601 catggcagaa gaatgggaat attttatct gacagaaatc agtaatatit atatattit
 10 661 attttaaaa tatttatta ttatttatt taagttcata ttccatattt attcaagatg
 721 ttttaccgta ataattatta ttaaaaatat gcttct (SEQ ID NO: 7)

GenBank Accession # XM_003751:

1 tctggaggat gtggctgcag agcctgtgc tcttgggac tgtggcctgc agcatctctg
 15 61 caccggccc ctcgccagc cccagcagc agccctggga gcatgtgaat gccatccagg
 121 aggccggcg tctctgaac ctgagttag agactgtgc tgagtgaat gaaacagtag
 181 aagtcattctc agaaatgttt gacctcagg agccgacctg cctacagacc cgccctggagc
 241 tgtacaagca gggcctgcgg ggcagcctca ccaagctcaa gggcccttg accatgatgg
 301 ccagccacta caagcagcac tgcctccaa ccccgaaac ttctgtgca acccagacta
 361 tcaccttga aagtttcaa gagaacctga aggactttct gcttgctc ccccttgact
 20 421 gctgggagcc agtccaggag tgagaccggc cagatgaggc tggccaagcc ggggagctgc
 481 tctctcatga aacaagagct agaaactcag gatgtgctc ttgaggggac caagggtg
 541 gccacagcca tgggtggagt ggcctggacc tgcctgggc cacactgacc ctgatacagg
 601 catggcagaa gaatgggaat attttatct gacagaaatc agtaatatit atatattit
 25 661 attttaaaa tatttatta ttatttatt taagttcata ttccatattt attcaagatg
 721 ttttaccgta ataattatta ttaaaaatat gcttct (SEQ ID NO: 8)

General Target Regions:

- 30 (1) 5' Untranslated Region - nts 1 - 32
 (2) 3' Untranslated Region - nts 468 - 789

Initial Specific Target Motif:

- Group I AU-Rich Element (ARE) Cluster in 3' untranslated region
 35 5' AUUUUUUUUUUUUUUUUUUU 3' (SEQ ID NO: 1)

5.3. Interleukin 2 ("IL-2")

GenBank Accession # U25676:

1 atactctct ttaactacta ctcacattaa cctcaactcc tgcacaaatg tacaggatgc
 5 61 aactcctgtc ttgcattgca ctaattcttg cactgtgtac aaacagtgtca cctacttcaa
 121 gtgcgacaaa gaaaacaaag aaaacacagc tacaactgga gcattactgc ctggatttac
 181 agatgatttt gaatggaatt aataattaca agaattccaa actcaccagg atgtccacat
 241 ttaagtttta catgcccaag aaggccacag aactgaaaca gcttcagttg ctagaagaag
 301 aactcaaac cttggaggaa gtgctgaatt tagctcaag caaaaacttt cactaagac
 10 361 ccaggcgact aatcagcaat atcaacgtaa tagttctgga actaaaggga tctgaacaaa
 421 cattcatgtg tgaatatgca gatgagacag caaccattgt agaatttctg aacagatgga
 481 ttaccttttg tcaaaagc atctcaacac taacttgata attaagtgtc toccacttaa
 541 aacatatcag gcccttattt tatttattta aatatllaaa ttltatattt attgttgaat
 601 gtatgggtgc tactattgt aactattatt cttaatttta aaactataaa tatggatctt
 15 661 ttatgattct ttttgaagc ctaggggctg ctaaaatggt ttacttatt tatcccaaaa
 721 atatttatta ttatgttga tgttaaatat agtatctatg tagattgggt agtlaaacta
 781 tttataaat ttgataaata taaaaaaaa aaacaaaaaa aaaaa (SEQ ID NO: 9)

General Target Regions:

- 20 (1) 5' Untranslated Region - nts 1 - 47
 (2) 3' Untranslated Region - nts 519- 825

Initial Specific Target Motifs:

- Group III AU-Rich Element (ARE) Cluster in 3' untranslated region
 25 5' NAUUUAUUUAUUUAN 3' (SEQ ID NO: 10)

5.4. Interleukin 6 ("IL-6")

GenBank Accession # NM_000600:

1 ttctgccctc gagccaccg ggaacgaaag agaagctcta tctgcctcc aggagcccag
 30 61 ctatgaactc cttctccaca agcgcttctg gtccagtgc cttctccctg gggctgtctc
 121 ttgtgtgtcc ttgtgccttc cctgcccag tacccccagg agaagattcc aaagatgtatg
 181 ccgcccacac cagacagcca ctcacctt cagaacgaat tgacaacaa attcgggtaca
 241 tctcgcacgg catctcagcc ctgagaagg agacatgtaa caagagttaac atgtgtgaaa
 301 gcagcaaaaga ggcactgtgca gaaaacaacc tgaaccttcc aaagatggct gaaaagatg
 35 361 gatgtctcca atctgtgatic aatgaggaga ctgctctggt gaaaatcatic actggtcttt
 421 tggagtttga ggtataccta gattacctcc agaacagatt tgagagtagt gaggaacaag

481 ccagagctgt gcagatgagt acaaaagtc tgatccagtt cctgcagaaa aaggcaaga
 541 atctagatgc aataaccacc cctgacccaa ccacaaatgc cagctgctgc acgaagctgc
 601 aggcacagaa ccagtgcctg caggacatga caactcatct catctgcgc agctttaagg
 5 661 agttcttga gtccagcctg agggctcttc ggcaaatgta gcatgggcac ctacatgtt
 721 tgttgttaat gggcattctc tctctggtc agaaacctgt ccactgggca cagaacttat
 781 gttgtictct atggagaact aaaagtatga gcgttaggac actatttaa ttattttaa
 841 ttattataa tttaaatag tgaagctgag ttaatttat taatcatat ttattttt
 901 aagaagtacc acttgaaca ttatgtat tagttttaa ataataatgg aaagtggcta
 10 961 tgcagttga atatccttg tticagagcc agatcatttc ttggaaatg taggcttacc
 1021 tcaataaat gctaaacta tacatattt taagaaata ttatatigt attatataa
 1081 tgtataatg gttttatac caataaatgg cattttaa aatic (SEQ ID NO: 11)

General Target Regions:

- 15 (1) 5' Untranslated Region - nts 1 - 62
 (2) 3' Untranslated Region - nts 699 - 1125

Initial Specific Target Motifs:

- 20 Group III AU-Rich Element (ARE) Cluster in 3' untranslated region
 5' NAUUUAUUUAUUUAN 3' (SEQ ID NO: 10)

5.5. Vascular Endothelial Growth Factor ("VEGF")

GenBank Accession # AF022375:

1 aagagctcca gagagaagtc gaggaagaga gagacggggt cagagagagc gcgcgggcgt
 25 61 gcgagcagcg aaagcgacag gggcaaatg agtgcactgc ttgggggt gaccgcggga
 121 gcgcggcgtg agccctccc ctgggatcc cgcagctgac cagtcgcgt gacgcacaga
 181 cagacagaca ccgccccag cccagttac cactctctc ccggccggcg gcggacagtg
 241 gacgcggcg cgagccgcgg gcagggggcg gagccggccc ccggaggcgg ggtggaggcg
 301 gtcggagctc gcggcgtgc actgaactt ttgtccaac ttctggctg ttctgcctc
 361 ggaggagcgg tggtcgcgc gggggaagcc gagccgagcg gagcccgag aagtcttagc
 421 tcgggcccgg aggagccgca gccggaggag ggggaggagg aagaagagaa ggaagaggag
 481 agggggccgc agtggcgact cggcgctcgg aagccggct catggacggg tgaggcggcg
 541 gtgtgcgag acagtgtctc agcgcgcgcg ctcccaagc ctggccggcg ctggccggcg
 601 gaggaagagt agctcgcga ggcgcggagg agagcgggcc gccccacagc ccgagccgga
 661 gagggacgcg agccgcgcgc cccgtcggg cctccgaac catgaactt ctgtgttt
 35 721 ggtgtcattg gagccttgc ttgtctctc acctccacca tgccaagtgg tccagcgctg

781 caccatggc agaaggagga gggcagaatc atcacgaagt ggtgaagttc atggatgct
 841 atcagcgag ctactgccat ccaatcgaga cctggtgga catcttcag gagtacctg
 901 atgagatga gtacatctc aagccatct gtgtgccct gatgcgatc gggggctgct
 5 961 ccaatgacga gggcctggag tgtgtgccc ctaggagatc caacatcacc atgcagatta
 1021 tgcggatcaa acctcaccan ggcagcaca taggagagat gagcttcta cagcacaaca
 1081 aatgtgaatg cagaccaaag aaagatagag caagacaaga aaatccctgt gggccttct
 1141 cagagcggag aaagcatttg ttgtacaag atccgcagac gtgtaaatgt tccctgcaaaa
 1201 acacacactc gcgttgcaag gcgaggcagc ttgagftaaa cgaacgtact tgcagatgtg
 10 1261 acaagccgag gcggtgacc gggcaggagg aaggagcctc cctcagggtt tcgggaacca
 1321 gatctctc caggaaagac tgatacagaa cgtatgatac agaaccacg ctgccgccac
 1381 cacacatca ccatgacag aacagtcctt aatccagaaa cctgaaatga aggaagagga
 1441 gactctgcgc agagcacttt gggtcggag ggcgagactc cggcggaagc attccggggc
 1501 gggtagacca gcacggtccc tcttgaatt ggatcgcca ttattttt ctgtgcta
 15 1561 aatcccgag ccgggaagat tagagagttt tattctggg attcctgtag acacaccac
 1621 ccacatcat acattatat atatatata tatatatata taaaataaa tatcttatt
 1681 tatatatat aaaatatata tatictttt taaattaac agtgcatag ttattggigt
 1741 ctctactgga tgtattgac tgcgtggac ttgagttgg aggggaatgt tccacticag
 1801 atcctgacag ggaagaggag gagatgagag actctggcat gatcttttt ttgtcccact
 1861 tgggtgggccc aggtctctct cccctgccc agaattgtca aggcaggggc atgggggcaa
 20 1921 atatgaccca gttttgggaa caccgacaaa cccagccctg gcgctgagcc tcctacccc
 1981 aggtcagacg gacagaaaga caaatcacag gtccgggat gaggacaccg gctctgacca
 2041 ggagtttggg gagcttcagg acattgctgt gctttgggga ttccctcac atgctgcacg
 2101 cgcattctgc cccaggggc actgcctgga agattcagga gcctgggcgg ccttcgctta
 2161 ctctacctg cttctgagtt gcccaggagg ccactggcag atgtcccgcc gaagagaaga
 22 2221 gacacattgt tggagaagc agcccatgac agcgcacctt cctgggactc gccctcatcc
 2281 tcttctgct cccctctct ggggtgcacc taaagagacc tatgtcctca caccattgaa
 2341 accactagtt ctgtccccc aggaaacctg gttgtgtgt tgtgagttgt tgaccttct
 2401 ccattccctg gtccttccct tccctcccg aggcacagag agacagggca ggaaccagct
 30 2461 gccattgtg gaggcagaga aaagagaang tgtttatat acggtactta tttaatatcc
 2521 ctttttaatt agaaattaga acagttaatt taattaaaga gtaggggttt tttagatg
 2581 tcttggttaa tatttaatt caactatffa tgagatgtat ctttgcct cctgtcct
 2641 ctatttgta ccggttttt tatataaat tcatgttcc aatctctc tccctgatcg
 2701 gtgacagtca ctagcttate ttgaacagat atttaattt gctaacactc agctctgccc
 2761 tcccgatcc cctggctccc cagcacat tctttgaaa gagggtttca atatacatct
 35 2821 acatacata tatatatagg gcaacttgta ttgtgtgta tatatatata tatatgtta

2881 tgtatatatg tgatcctgaa aaaataaaca tcgtattct gtittttata tgttcaaacc
 2941 aaacaagaaa aatagagaa ttctacatac taatctctc tccittttta attttaatat
 3001 ttgttatcat ttatttatg gtgctactgt ttatccgtaa taattgtggg gaaaagatat
 5 3061 taacatacag tctttgtctc tagtgcagti ttccgagata ttccgtagta catatttatt
 3121 tttaacaac gacaaagaaa tacagafata tcttaaaaaa aaaaa (SEQ ID NO: 12)

General Target Regions:

- (1) 5' Untranslated Region - nts 1 - 701
- 10 (2) 3' Untranslated Region - nts 1275 - 3166

Initial Specific Target Motifs:

- (1) Internal Ribosome Entry Site (IRES) in 5' untranslated region nts 513 -704
 5'CCGGGCUCAUGGACGGGUGAGGCGGCGUGUGCGCAGACAGUG
 CUCCAGCGCGCGCGCUCGCCAGCCCGGCCGCGCCGCGCGG
 15 AGGAAGAGUAGCUCGCGGAGGCGCCGAGGAGAGCGGGCCGCCCC
 ACAGCCCGAGCCGAGAGGGACGCGAGCCGCGCGCCCGGUCGG
 GCCUCCGAAACCAUGAACUUUCUGCUGUCUUGGGUGCAUUGGAG
 CCUUGCCUUGCUGCUCUACCUCCACCAUG 3' (SEQ ID NO: 13)
- 20 (2) Group III AU-Rich Element (ARE) Cluster in 3' untranslated region
 5' NAUUUAUUUAUUUAN 3' (SEQ ID NO: 10)

5.6. Human Immunodeficiency Virus I ("HIV-1")

GenBank Accession # NC_001802:

1 ggctctctg gttagaccag atctgagcct gggagctctc tggctaacta gggaaacccac
 25 61 tgcttaagcc tcaataaagc ttgccttgag tgcttcaagt agtgtgtgcc cgtctgttgt
 121 gtagctctgg taactagaga tccttcagac ccttttagtc agtgttgaaa atctctagca
 181 glggcgcccg aacaggggacc tgaagcgaa agggaaacca gaggagctct ctcgacgcag
 241 gactcggcct gctgaagcgc gcacggcaag aggcgagggg cggcgactgg tgagtacgcc
 30 301 aaaaattttg actagcggag gctagaagga gagagatggg tgcgagagcg tcagtattaa
 361 gggggggaga attagatcga tgggaaaaaa ttcggfttaag gccaggggga aagaaaaaat
 421 ataaattaaa acatatagta tgggcaagca gggagctaga acgattcgca gttaatcttg
 481 gcctgttaga aacatcagaa ggcgtgtagc aaatactggg acagctacaa ccatcccttc
 541 agacaggatc agaagaactt agatcattat ataatacagt agcaaccctc tatgtgtgtg
 601 atcaaaggat agagataaaa gacaccaagg aagctttaga caagatagag gaagagcaaa
 35 661 acaaaagtaa gaaaaaagca cagcaagcag cagctgacac aggcacacagc aatcaggtca

721 gccaaaatta cccatagtg cagaacatcc aggggcaaat gggtacatcag gccatcatcac
 781 ctagaacttt aaatgcatgg gtaaaagtag tagaagagaa ggccttcagc ccagaagtga
 841 tacccatgtt ttcagcatta tcagaaggag ccaccccaca agatttfaac accatgctaa
 5 901 acacagtggg gggacatcaa gcagccatgc aaatgttaaa agagaccatc aatgaggaag
 961 ctgcagaatg ggatagagtgc catccagtcg atgcagggcc tattgcacca ggcagatga
 1021 gagaaccaag gggaaatgac atagcaggaa ctactagtac ccttcaggaa caaataggat
 1081 ggaigacaaa taatccacct atcccagtag gagaaaltta taaaagatgg ataactctgg
 1141 gattaaataa aatagtaaga atgtatagcc ctaccagcat tctggacata agacaaggac
 1201 caaaggaacc ctttagagac tatgtagacc ggttctataa aactctaaag gccagagcaag
 1261 cttcacagga ggtaaaaaat tggatgacag aaactctgtt ggtccaaaat gcgaaccag
 1321 atgtgaagac tattttaaaa gcattgggac cagcggctac actagaagaa atgatgacag
 1381 catgtcaggg agtaggagga cccggccata aggcaagagt ttggctgaa gcaatgagcc
 1441 aagtaacaaa ttcagctacc ataagtatgc agagaggcaa ttttagaac caaagaaga
 1501 ttgttaagtg ttcaatgtg gccaagaag ggcacacagc cagaattgc agggccccta
 1561 ggaaaaaggg ctgttggaag ttgtgaaagg aaggacacca aatgaagat tgtacttgaga
 1621 gacaggctaa tttttaggg aagatctggc ctctctcaa gggaaggcca gggaatttc
 1681 ttcagagcag accagagcca acagcccac cagaagagag cttcaggctt gggtagagga
 1741 caacaactcc cctcagaag caggagccga tagacaagga actgtatcct ttaactccc
 1801 tcaggctact ctttggcaac gacccctcgt cacaataaag ataggggggc aactaaagga
 1861 agctctatta gatacaggag cagatgatac agtattagaa gaaatgagtt tgccagggaag
 1921 atggaaacca aaaatgatag ggggaattgg aggttttate aaagtaagac agtatgatca
 1981 gatactcata gaaatctgtg gacataaagc tataggtaca gtattagtag gacttacacc
 2041 tgicaacata attggaagaa atctgtgac tcagattggt tgcactttaa attttccat
 2101 tagccctatt gagactgtac cagtaaaatt aaagccagga atggatggcc caaaagttaa
 2161 acaatggcca ttgacagaag aaaaaataaa agcattagta gaaatttga cagagatgga
 2221 aaagggaagg aaaatttcaa aaattgggcc tgaataacca tacaatactc cagtatttgc
 2281 cataagaaaa aaagacagta ctaaatggag aaatagta gatctcagag aacttaataa
 2341 gagaactcaa gactctggg aagticaatt aggaatacca catcccgag ggttaaaaaa
 2401 gaaaaaatca gtaacagtac tggatggggg tgaatcatat ttacagtc ccttagatga
 2461 agacttcagg aagtatactg catctaccat acctagtata aacaatgaga caccagggat
 2521 tagatatcag tacaatgtgc ttccacaggg atggaagga tcaccagcaa tattccaag
 2581 tagcatgaca aaaatcttag agccctttag aaacaaaaat ccagacatag ttatctatca
 2641 atacatggat gatttgtatg taggatctga cttagaataa gggcagcata gaacaaaaat
 2701 agaggagctg agacaacatc ttgttaggtg gggacttacc acaccagaca aaaaacatca
 2761 gaaagaacct ccattccttt ggaatgggta tgaactccat cctgataaat ggacagtaca

2821 gcctatagtg ctgccagaaa aagacagctg gactgticaat gacalacaga agttagtggg
 2881 gaaatigaat tgggcaagtc agatttacc agggattaaa gtaaggcaat tatgtaaact
 2941 ccttagagga accaaagcac taacagaagt aataccacta acagaagaag cagagctaga
 5 3001 actggcagaa aacagagaga ttctaaaaga accagtacat ggaagtgtatt atgacccatc
 3061 aaaaagacta atagcagaaa tacagaagca ggggcaaggc caatggacat atcaaatita
 3121 tcaagagcca ttaaaaatc tgaacaagg aaaatatgca agaattgagg gtgcccacac
 3181 taatgatgta aaacaattaa cagaggcagt gcaaaaaata accacagaaa gcatagtaaf
 3241 atggggaaag actcctaaat ttaactgcc catacaaaag gaacatggg aaacatggtg
 10 3301 gcagagtgat tggcaagcca cctggatcc tgaaggagg tttgttaata cccctccct
 3361 agtgaatta tgggtaccagt tagaagaaga acccatagta ggaagcagaaa ccttctatgt
 3421 agatggggca gctaacagg agactaaatt aggaagaaga ggatatgta ctaatagagg
 3481 aagacaaaaa gttgtcacc taacigacac aacaatcag aagactgagt tacaagcaat
 3541 ttatctagct ttgcaggatt cgggattaga agtaaacata gtaacagact cacaatatgc
 15 3601 attaggaatc attcaagcac aaccagatca aagtgaatca gagttagtca atcaaatat
 3661 agagcagtha ataaaaaagg aaaaggtcta tctggcatgg gtaccagcac acaaaagaa
 3721 tggaggaaat gaacaagtag ataaattagt cagtctgga atcaggaaag tactatttt
 3781 agatggaata gataaggccc aagatgaaca tgagaatat catagtaatt ggagagcaat
 3841 ggctagtgat ttaaccigc cactgtagt agcaaaagaa atagtagcca gctgtgataa
 20 3901 atgtcagcta aaaggagaag ccatgcatgg acaagtagac tgtagtcagg gaatatggca
 3961 actagattgt acacatttag aaggaaaaat tatcttgta gcagttcatg tagccagtgg
 4021 atatatagaa gcagaagtha ttccagcaga aacagggcag gaaacagcat attttcttt
 4081 aaaattagca ggaagatggc cagtaaaac aatacatct gacaatggca gcaatttca
 4141 cgggtctacg gttaggggccg cctgttggg ggcgggaatc aagcaggaa tttggaattcc
 25 4201 ctacaatccc caaagtcaag gagttagta atctatgaat aaagaattaa agaaattat
 4261 aggcaggta agagatcagg ctgaacatct taagacagca gtacaaatgg cagtattcat
 4321 ccacaatttt aaaagaaaag gggggattgg ggggtacagt gcaggggaaa gaatagtaga
 4381 cataatgca acagacatac aaactaaaga attacaaaa caaattacaa aaattcaaaa
 4441 ttctgggtt tattacaggc acagcagaaa tccactttgg aaaggaccag caaagctct
 30 4501 ctggaagggt gaaggggcag tagtaataca agataatagt gacataaaag tagtgcacag
 4561 aagaaagca aagatcatta gggattatgg aaacagatg gcagggtgat attgtgtggc
 4621 aagtagacag gatgaggatt agaacaatgga aaagtttagt aaacaccat atgtatgttt
 4681 cagggaagc taggggatgg tttatagac atcactatga aagccctcat ccaagaataa
 4741 gttcagaagt acacatccca ctagggatg ctagattgtt aataacaaca tatggggc
 35 4801 tgcatacagg agaaagagac tggcatttgg gtcagggatg ctccatagaa tggagaaaa
 4861 agagatatag cacacaagta gaccctgaac tagcagacca actaatcat ctgtattact

4921 ttgactgttt ttgactct gctataagaa aggccttatt aggacacata gttgcccata
 4981 ggtgtgaata tcaagcagga catacaagg taggatctct acaactctg gcactagcag
 5041 catlaataac accaaaaaag ataaagccac ctctgcctag tgttacgaaa ctgacagagg
 5 5101 atagatggaa caagcccccag aagaccaagg gccacagagg gagccacata atgaatggac
 5161 actagagctt ttgaggagc ttaagaatga agctgttaga catttcccta ggaattggct
 5221 ccatgcttta gggcaacata tctatgaac ttatggggat acttggcgag gagtggaaac
 5281 cataataaga attctgaac aactgtgtt tatccattt cagaattggg tctcgacata
 5341 gcagaatagg cgttactcga cagaggagag caagaaatgg agccagtaga tctagacta
 10 5401 gagccctgga agcatccagg aagtcagcct aaaactgctt gtaccaatg ctatigttaa
 5461 aagtgttct ttcatgccca agttgtttc ataacaaaag ccttaggcac ctctatggc
 5521 aggaagaagc ggagacagcg acgaagagct catcagaaca gtcagactca tcaagctct
 5581 ctatcaaac agtaagtagt acatgtaat caacctatac caatagtagc aatagttaga
 5641 ttagttagt caataataat agcaatagtt gtgtgttcca tagtaatcat agaataatgg
 15 5701 aaaaatttaa gacaagaaa aatagacagg ttaattgata gactaataga aagagcagaa
 5761 gacagtggca atgagatgga agggagaata tcagcacttg tggagatggg ggtggagatg
 5821 gggcaccatg ctcttggga tgttgaatg ctgtatgct acagaaaaat tgtgggtcac
 5881 agtctattat ggggtacctg tgtggaagga agcaaccacc actctattt gtgcatacga
 5941 tgcataagca tatgatacag aggtacataa tgtttggcc acacatgcct gtgtacccac
 20 6001 agacccaac ccacaagaag tagtatgtt aaatgtgaca gaaaatttta acatgtggaa
 6061 aaatgacatg tagaacaga tgcattagga tataatcagt ttatgggac aaagccataa
 6121 gccatgtgta aaattaaccc cactctgtt tagtttaaag tgcactgatt tgaagaatga
 6181 tactaatacc aatagtagta gctggagaat gataatggag aaaggagaga taaaaaactg
 6241 ctcttcaat atcagcaca gcataagagg taagggtcag aaagaatag catttttta
 25 6301 taaacttgat ataatacaca tagataatga tactaccagc tataagtga caagtgttaa
 6361 cacctcagtc attacacagg cctgtccaaa ggtatcctt gagccaatic ccatacatta
 6421 ttgtgccccg gctgttttg cgattctaaa atgtaataat aagacgttca atggaacagg
 6481 accatgtaca aatgcagca cagtacaagt tacacatgga attaggccag tagtatcaac
 6541 tcaactgctg ttaaatggca gtctagcaga agaagaggta gtaatagat ctgtcaattt
 30 6601 caccggacaat gctaaaacca taatagtaca gctgaacaca tctgtagaaa ttaattgtac
 6661 aagacccaac aacaatacaa gaaaagaat ccgtatccag agaggaccag ggagagcatt
 6721 tttataata ggaaaatag gaaatfatag acaagcacat tgtacatta gttagacaaa
 6781 atggaataac actttaaac agatagctag caaattaa gaacaattg gaaataataa
 6841 aacaataac ttttaagcaat cctcaggagg ggacccagaa attgtaacgc acagtittaa
 35 6901 ttgtggaggg gaattttct actgtaatic aacacaactg ttaatagta ctgtgttaa
 6961 tagtacttgg agtactgaag ggtcaataaa cactgaagga agtgacacaa tcacctccc

7021 atgcagaata aaacaaatta taaacatgtc gcagaaaagta ggaaaagcaa tgtatgcccc
 7081 tcccatcagt ggacaaatta gatgttcac aaatattaca gggctgctat taacaagaga
 7141 tgggtgtaat agcaacaatg agtcgcagat cticagacct ggaggaggag atatgaggga
 5 7201 caattggaga agtgaattat ataaatataa agtagtaaaa attgaacct taggagtagc
 7261 accccaag gcaaaagaga gagtgtgca gagagaaaaa agagcagtgg gaataggagc
 7321 ttgttcctt gggttcttgg gaggcagcgg aagcactatg ggcgcagcct caatgacgct
 7381 gagggtacag gccagacaat taltgtctgg tatagtgcag cagcagaaca atttgcgtgag
 7441 ggcatttgag gcgcaacagc atctgttcca actcacagtc tggggcatca agcagctcca
 10 7501 ggcaagaatc ctggctgtgg aaagatacct aaaggatcaa cagctcctgg ggaattgggg
 7561 ttgctctgga aaactcatt gcaccatgc tgtgccttgg aatgctagtt ggagtataa
 7621 atctctggaa cagatttggg atcacacagc ctggatggag tgggacagag aaattaacaa
 7681 ttacacaagc ttaatacact ccttaatga agaatcgaa aaccagcaag aaaagaatga
 7741 acaagaatta ttggaattg ataatgggc aagtttggg aattggtta acataacaaa
 15 7801 ttggctgtgg tatataaaat tatcataat gatagtagga ggctgtgag gttaagaat
 7861 agtttttct gtactttcta tagtgaatag agttaggcag ggatattcac cattatcgtt
 7921 tcagaccac ccaccaacc cgaggggacc cgacaggccc gaaggaaatg aagaagaagg
 7981 tggagagaga gacagagaca gatccattcg attagtgaac ggatccttgg cacttatctg
 8041 ggacgatctg cggagcctgt gcctcttcag ctaccaccgc ttgagagact tactcttgat
 20 8101 tgtaacgagg attgtggaac ttctgggacg caggggggtg gaagccctca aatatgttg
 8161 gaatctcta cagtattgga gtcaggaaat aaagaatagt gctgttagct tgcctaattc
 8221 cacagccata gcagttagctg aggggacaga taggggtata gaagtagtac aaggagcttg
 8281 tagagctatt gccacatac ctagaagaat aagacagggc ttgaaagga ttitgctata
 8341 agatgggtgg caagtgggca aaaagtagtg tgaattgatg gcctactgta agggaaagaa
 25 8401 tgagcagcag tgagccagca gcagataggg tgggagcagc atctcagagc ctggaaaaac
 8461 atggagcaat cacaagtagc aatacagcag ctaccaatgc tgccttgccc tggctagaag
 8521 caaagagga ggaggagggt ggttttcag tcacacctca ggtacctta agaccaatga
 8581 cttaagaagc agctgtagat cttagccact tttaaaaaga aaaggggggc ctggaaggcc
 8641 taattcactc ccaagaaga caagatatcc ttgatctgtg gatctaccac acacaagct
 30 8701 attccctga ttacgagaac tacacaccag ggccaggggg agatattcca ctgaccttg
 8761 gatgtgtcta caagctagta ccagttgagc cagataagat agaagaggcc aataaaggag
 8821 agaacaccag ctgtttacac cctgtgagcc tgcattggat ggtgaccgc gagagagaag
 8881 ttttagagtg gaggtttgac agccgcctag cattcatca cgtggcccga gagctgcatc
 8941 cggagtactt caagaactgc tgacatcgag ctgtctacaa gggactttcc gctggggact
 9001 ttccagggag gcgtggcctg ggcgggactg gggagtgggc agccctcaga tctctcatat
 35 9061 aagcagctgc ttttgcctg tactgggtct ctctggttag accagatctg agcctgggag

9121 ctctctggct aactaggga cccactgctt aagcctcaat aaagcttgc ttgagtgcct
 9181 c (SEQ ID NO: 14)

5 Initial Specific Target Motifs:

- (1) Trans-activation response region/Tat protein binding site - TAR RNA - nts 1 - 60
 "Minimal" TAR RNA element
 5' GGCAGAUUCUGAGCCUGGGAGCUCUCUGCC 3' (SEQ ID NO: 15)
- 10 (2) Gag/Pol Frameshifting Site - "Minimal" frameshifting element
 5' UUUUUUAGGGAAGAUCUGCCUCCUACAAGGGAAGGCCAGG
 GAAUUUUCUU 3' (SEQ ID NO: 16)

5.7. Hepatitis C Virus ("HCV" - Genotypes 1a & 1b)

15 GenBank Accession # NC_001433:

1 ttggggcgca cactocacca tagatcaclc cctgtgagg aactactgtc ttacgcagga
 61 aagcgtctag ccatggcgctt agtatgagt ttgtgcagcc tccaggaccc cccctcccg
 121 gagagccata gtgtctctgc gaaccgtga gtacaccgga attgccagga gcacggggtc
 181 cttcttga tcaaccgct caatgcctgg agatttggc gtgccccgc gagactgcta
 241 gccagtagt gtgggtcgc gaaaggcctt gtgtactgc ctgatagggt gcttcgagt
 301 gccccgggag gtctctaga ccgtgcacca tgagcacaaa tctaaacct caaagaaaa
 361 ccaaacgtaa caccaaccgc cggccacagg acgttaagtt cccggcggtt ggtcagatgc
 421 ttgttgaggt ttacctgttg ccggcaggg gccccaggtt ggtgtgtcgc gcgactagga
 481 agacttccga gcggtcgcaa cctcgtggaa ggcgacaacc tatcccaag gctcggcgc
 541 ccgagggtag gactcgggt cagcccggtt acccttgcc ccttatggc aacgagggt
 601 tgggtgggc aggatggct ctgtacccc gtggtctctg gctatgttg ggcacacag
 661 accccggcg taggtgcgt aatttggga aggtatcga tacccttaca tgcggtctg
 721 ccgacctcat ggggtacatt ccgtgtgtg ggcggccct agggggcgtt gccagggcc
 781 tggcacatgg tgcggggt ctggaggacg gcgtgaacta tgcaacagg aatctgccg
 841 gttgtcttt ctctatctc ctctagctt tctgtcttg ttgaccate ccagttccg
 901 cttagcaggt gcgcaacgtg tccgggatat accatgtcac gaacgactgc tcaactcaa
 961 gtattgtgta tgaggcagc gacatgata tgcacacc cgggtgcgtg ccdgcgtcc
 1021 gggagagtaa ttctccgt tctgggtgag cgtcactcc cagctcgcg gccaggaaca
 1081 gcagcatccc caccagaca atacgagcc acgtcgattt gctcgttgg gcgctcgtc
 1141 tctgtccgc tatgtacgtt ggggatctct gcggatcgt ttctcgtc tccagctgt
 1201 teactcttc acctcggcg tatgagacgg tacaagattt caattgctca atctatccc

1261 gccacgtaic aggtcaccgc atggcttggg atatgatgat gaactgggtca cctacaacgg
 1321 cccatgtgggt atcgacgcta ctccggatcc cacaagccgt cgtggacatg gtggcggggg
 1381 cccactgggg tgtcctagcg ggcccttgct actattccat ggtgggggaa tgggctaagg
 5 1441 tctgtattgt gatgtacttc ttgtctggcg ttgacgggca caccacgtg acaggggggaa
 1501 gggtagcctc cagcacccag agcctcgtgt cctggctctc acaaggccca tctcagaaaa
 1561 tccaactcgt gaacaccaac ggcagctggc acatcaacag gaccgctctg aattgcaatg
 1621 actccctcca aactgggttc atfgctgcgc tgfictacgc acacagggtc aacggtccg
 1681 ggtgcccaga gcgcatggct agctgccgcc ccatcgatga gttcgctcag ggggtgggtc
 1741 ccatcactca tgatatgcct gagagctcgg accagaggcc atattgtcg cactacgcgc
 10 1801 ctgaccctg cgggatcgtg cctgcgtcgc aggtgtgtgg tccagtgtat tgcctactc
 1861 cgagccctgt tgtatggggg acgaccgacg gtttcggcgc tctactgtat agtggggggg
 1921 agaatgagac agacgtgctg ctacttagca acacggcgcc gcctcaaggc aactggtttg
 1981 ggtgcacgtg gatgaacagc actgggttca ccaagacgtg cggggggccct ccgtcaaca
 15 2041 tcgggggggt cggaacaac acctgggtct gcccacgga ttgcttcgg aagcaccccg
 2101 agggcactta cacaaagtgt ggctcggggc cctgggtgac acccaggtgc atgggttgact
 2161 acccatagag gctctggcac tacccttgca ctgtaactt taccgtctt aaggtcagga
 2221 tgtatgtggg gggcgtggag cacaggtca atgctgcatg caattggact cgagagagagc
 2281 gctgtgactt gggagacagg gataggctag aactagccc gctgctgctg tctacaacag
 2341 agtggcagat actgccctgt tcttcacca cctaccgcg cctgtccact ggcctgacc
 20 2401 atcttcaccg gaacatcgtg gacgtgcaat acctgtacgg tatagggtcg gcagttgtct
 2461 cctttgcaat caaatgggag tatatcctgt tgccttctct tctctggcg gacgcgcgcg
 2521 tctgtgcctg cttgtggatg atgctgctga tagcccagcg tgaggccacc ttagaagacc
 2581 tgggtgtcct caatggcgcg tctgtggccg gaggcgcctg ccttctctcc ttctctgtgt
 2641 tcttctgcgc cgccctgtac atcaaaaggca ggctggctcc tggggcgggca tatgtctct
 2701 atggcgtatg gccgttgctc ctgctcttgc tggccttacc acccagagct tatgccatgg
 2761 accgagagat ggctgcatcg tgcggaggcg cgtttttgt aggtcttgta cttctgacct
 2821 tgcaccata ctataaggfg ttctcgtcta ggctcatatg gtggttaca tatattata
 2881 ccagagccga ggcgcacttg caagtgtggg tccccctct caatgttcgg gggagccgcg
 2941 atgcacatcat cctcttaca tgcgggtcc atccagagct aatctttgac ataccaaac
 3001 tctcgtcgc cactacgggt ccgctcatgg tgcctcagcg tgcataact agagtggcgt
 3061 actttgtacg cgctcagggg ctcacccgtg catgcatgtt agtgcggaag gtcgtggag
 3121 gccactatgt ccaaatggcc ttcatgaagc tggccgcgct gacaggtacg tacgtatatg
 3181 accactttac tccactcggg gatggggccc acgggggcct acgagacctt gcggtggcag
 3241 tagagcccggt cgtcttctct gacatggaga cttaactcat cacctggggg gcagacaccg
 35 3301 cggcgtgtgg ggacatcatic tgggtcttac cagtctccgc ccgaaggggg aaggagatac

3361 ttctaggacc ggccgatatg ttggagagc aggggtggcg gctcctfgcg cctatcagcg
 3421 cctattccca acaaacggcg ggccgtcttg gctgtatcat cactagcctc acaggctggg
 3481 acaagaacca ggtcgaatgg gagggttcagg tgctctccac cgcaacgcaa tcttctctgg
 5 3541 cgacctgcgt caatggcggtg tgttggaccg tctaccatgg tgccggctcg aagacctctg
 3601 ccggcccgaa gggtccaate accaanaatg acaccaatgt agaccaggac ctctgtggct
 3661 ggccggcgcc ccccggggcg cgtccatga caccgtgcac ctgcggcagc tggagacttt
 3721 acttggtcac gaggcatgct gatgtcgttc cgggtcgccg gcggggcgac agcaggggga
 3781 gctcgttttc cccaggcccc atctctacc tgaaggctc ctggggtgga ccaactgttt
 10 3841 gcccctggg gcacgttgtg ggcactttcc gggtcgtgt gtgcacccgg ggggttgcga
 3901 agggggtgga ctctataccc gttgagtcta tggaaactac catcggtct ccggtctca
 3961 gagacaactc atcccctcgg gcgctaccgc aaacttcca agtggcacat ttacacgtc
 4021 ccactggcag cggaagagc accaaagtgc cggctgcata tgcagcccaa ggttacaagg
 4081 tgcctgtcct aaaccctgcc gttgccga cattgggtt tggagcgtat atgtccaagg
 15 4141 cacatggcat cgagcctaac atcagaactg gggttaaggac catcaccacg gcggcccca
 4201 tcactgtact cacctatlgc aagtctctg ccgacggtag atgtcccggg ggcgcctatg
 4261 acatcataat atgtgatgaa tggcactcaa ctgactcgac taccatctg ggcactggca
 4321 tactcttggg taaggcagag acggtctggag cgcggctcgt cgtgtctgcc accgccacg
 4381 ctccgggac gatcacctg ccacaccca acatcgagg agtggccctg tccaacactg
 4441 gagagattcc ctctatggc aaagccatcc cacttgggc catcaagggg ggaaggcact
 20 4501 tcattctctg ccattccaag aagaagtgtg acgactcgc cgcaaaagtg acaggcctc
 4561 gactcaatgc ttagcgtgat taccgggtgc tcatgtgtc gtcataccg actagcggag
 4621 acgtcgttgt cgtggcaaca gacgtcttaa tgcgggttt taccggcgac ttgtactcag
 4681 tcatcactg caacacatgt gtcaccaga cagtctgatt cagctggat ccaacttca
 25 4741 cacttggagc gacaacgtg cccaagacg cgtgtctcgc tgcgacggc cgaggttaga
 4801 ctggcagggg caggagtggc atctacaggt ttgtgactcc aggaagacgg cctcaggca
 4861 ttttgactc ctgcgtctg tgtgagtgt atgacgcagg ctgccttgg tatgactca
 4921 cgcccgctga gacctcgggt aggttgcggg cttaactaaa tacaccaggg ttgccgtct
 4981 gccaggacca cctagatgic tgggagagcg tcttcacagg cctcaccac atagatccc
 30 5041 acttctgtc ccagaccana caggcaggag acaacctccc ctacttgga gcataccaag
 5101 ccacagtgtg gcgcagggct caggctccac ctccatcgtg ggaccaaatg tgaagtgtc
 5161 tcatacggct aaagcccaca ctgcattggc caacgccct cgtgtacagg ctaggagccg
 5221 ttcaaatga ggtcactctc acacaccca taaccaata catcatggca tgcattcgg
 5281 ctgacctgga ggtcgtcact agcacctggg tgcgtatagg cggagtctt ggcgtctgg
 5341 ccgctactg cctgacgaca ggacgcgtgg tcaattgtgg caggatcatc ttgtccggga
 35 5401 ggocagctgt tatcccgac aggggaagtc tctaccaga gticgatgag atggaagagt

5461 gtgtcttaca cctcccttac atcgagcaag gaatgcagct cggcagacaa ttcaaacaga
 5521 aggcgtctgg attgtctcaa acagccacca agcaagcggg ggctgtgctt cccgtgtgtg
 5581 agtccaagtg gcgagccctt gaggctctct gggcgaaaca catgtggaac ttcatcagcg
 5641 ggatacagta ctggcgagcg ctatccactc tggcgggaaa ccccgcgata gcatcattga
 5701 tggcttttac agcctctatc accagcccg caccaccaaa aataccctc ctgtttaaca
 5761 tcttgggggg atgggtggct gcccgaactc gctccccag cgtctcttcg gcttctgtgg
 5821 gcgcggcgat tgcgggtgcg gccgttgcca gcataggtct cgggaaggta ctgtgggaca
 5881 ttctggcggg ctatggggcg ggggtggctg gcgcactcgt ggcctttaa gtcattgagcg
 5941 gcgagatgcc ctccactgag gatctgggta attactccc tggcatcctt tctctggcg
 6001 cctgtgtgtt cggggtcgtg tgcgcagcaa tactgcgtcg gcacgtggcg cgggagaggg
 6061 gggctgtgca gtggatgaac cggctgatag cgttcgcttc ggggggtaac cagctctccc
 6121 ccacgacta tgtcccagag agcgacgccc cggcgcgctgt tactcagatc ctctccagcc
 6181 ttaccatcac tcagttgctg aagaggcttc atcagtggaat taatgagac tgcctcagc
 6241 ctgttccgg ctcgtggcta aaggatgttt gggactggat atgcacggtg ttgagtgaat
 6301 tcaagacttg gctccagtc aagctcctcg gcggtgtacc gggactccct ttctgtcat
 6361 gccaacgccc gtacaaggga gtctggcggg gggatggcat catgcaaac acctgccat
 6421 gtggagcaca gatcaccgga catgtcaaaa atggctccat gaggattgtt gggcacaana
 6481 cctgcagcaa cacgtggcat ggaacatfcc ccatcaacgc ataccacac gggcccttga
 6541 cgcctcccc agcgccgaac tattccaggg cgtctgtgcg ggtggtcgtt gaggagtacg
 6601 tggaggttac ggggtgggg gatttccact acgtgacggg catgaccact gacaacgtga
 6661 aatgcccatt ccaggttcca gccctggaat tticacgga ggtgagtgga gtacggttgc
 6721 acaggtatgc tccagtgtgc aaacctctcc tacgagagga ggtcgtatc caggtcgggc
 6781 tcaaccagta cctggtcggg tcacagctcc catgtgagcc cgaaccggaat gtggcagtcg
 6841 tcaattccat gtcaccgac cctctcata ttacagcaga gacggccaag cgtaggtctg
 6901 ccagggggtc tccccctcc ttggccagct cticagtag ccagttgtt gcgccttct
 6961 tgaaggcag atgtactacc catcatgact ccccgagcgc tgacctcaco gaggccaacc
 7021 tctgtggcg gcagagagtg ggcgggaaca tcaccgtgtt ggaatcagaa aataagggtg
 7081 taatcttga ctcttctcat ccgattcggg cgggtgagga tgagagggaa atatcgttc
 7141 cggcgagat cctgcgaaaa cccaggaagt tccccacgc gttgccata tgggcacgcc
 7201 cggattacaa cctctactg cttaggtctt ggaagggacc ggaactagtc ccccggttgg
 7261 tacacgggtg ccttttgcca tctaccaagg ccccccaat accacctcca cggaggaaga
 7321 ggaagggtgt cctgacagag tccacctgtt ctctgcctt ggcggagctc gctactaaga
 7381 cctttggcag ctccgggtcg tggccgttg acagcggcac ggcgactgco cctcccgatc
 7441 aggcctccga cgaaggcgac aaaggatccg acgttgagtc gtactctctc atgccccccc
 7501 tcgagggaga gcagggggac ccgacctca gcgacgggtc ttgttctacc gtgagcgggg

7561 aagctggtga ggacgtctgc tctgtctcaa tctctatac atggacaggt gccctgatca
 7621 cgccatgcgc tgcggaggag agcaagttgc ccatcaatcc gttagacaac tcttctgtgc
 7681 gtaccacagc tatggtctac tccacaacat ctgcagcgc aagtctgcgg cagaagaagg
 5 7741 tcactttga cagactgcaa gtctggagc accactaccg ggacgtctgc aaggagatga
 7801 aggcgaaggc gtccacagtt aaggctagcc ttctatctat agaggaggcc tgcnaactga
 7861 cgcgccca ttcggccaaa tccaatttg ctacggggc gaaggacgtc cggagcctat
 7921 ccagcagggc cgtcaaccac atccgtctcg ttgtggagga ctgtctgaa gacactgaaa
 7981 caccaattga taccaccatc atggcaaaaa atgagggttt ctgcgtcaa ccagagaagg
 8041 gaggcgcaa gccagctcgc ctatctgtat tccagacct gggggtagct gtatcgagaa
 10 8101 agatggccct ttacgacgtg gtctccacc ttctcaggc cgtgatgggc ccctcatacg
 8161 gatccagta ctctctggg cagcgggctg agttctggt gaatactgg aaataaaga
 8221 aatgccctat gggcttctca tatgacacc gctgottga ctcaacgctc actgagaatg
 8281 acatccgtac tgaggaaatc atttacaat gtgtgactt ggccccgaa gccaggcagg
 8341 ccataaggct gctcacagag cggctttatg tcgggggtcc ctgactaat tcgaaggggc
 15 8401 agaaactcgg ttatgcggg tgccgcgcaa gtggcgtgct gacgactagc tcggccaaca
 8461 cctcacatg ttacttgaag gccactcggc cctgtcgagc tgcaaatgc caggactgca
 8521 cgatgtcgt gaacggagac gacttgtc ttatctgtga gagtgcggga acccaggagg
 8581 atgcggcggc cctacgagcc ttacggagg ctagactag gtattccgc cccccgggg
 8641 acccgccca accagaatac gacttggagc tgataacgtc atgtctctcc aatgtctcgg
 20 8701 tcggcacaga tgcacccgc aaaagggtgt actacctcac ccgtgacccc accaccccc
 8761 tcgcacgggc tgcgtggag acagtttagc acactccagt caactctgg ctagccaata
 8821 tcatcatgta tgcgccacc ctatggcga ggtatgtct gatgactcat ttcttctta
 8881 tctcttagc tcaggagcaa ctfgaaaaag ccctgattg tcagatctac ggggcctgtt
 8941 actccattga gccacttgac ctacctcaga tcattgaac actccatgt cttagcgcat
 25 9001 ttctactca cagttactct ccaggtgaga tcaatagggt ggcttcatgc ctacggaac
 9061 ttggggtaac gcccttgcga gtctggagac atcggggcag aagtgtccg ctaagctac
 9121 tgtccaggg ggggagggt gccacttgcg gcaagtaact cttaacttg ccagtaaaaga
 9181 ccaagcttaa actcactcca atcccggctg cgtccagct agactgttc ggctgtgttc
 9241 ttgtgtgta caacggggga gacatatata acagcctgtc tctgccccga cccgttgtt
 30 9301 tcatgtgtg cctactcta ctttctgtag ggttaggcat ctacctgtc cccaaccgtt
 9361 gaacggggg ctaaccacto caggccaata ggccattccc tttttttt ttc (SEQ ID NO: 17)

General Target Region:

35 5' Untranslated Region - nts 1 - 328 - Internal Ribosome Entry Site (IRES):

5'UUGGGGGCGACACUCCACCAUAGAUCACUCCCCUGUGAGGAACUACUGUCUU
 CACGCAGAAAGCGUCUAGCCAUGGCGUUAUGUAGUGUGUGCAGCCUCCA
 GGACCCCCCUCGCGGAGAGCCAUAUGUGGUCUGCGGAACCGGUGAGUACACC
 5 GGAUUUGCCAGGACGACCGGGUCCUUUCUUGGAUCAAACCGCUCAAUGCCUGG
 AGAUUUGGGCGUGCCCCCGGAGACUGCUAGCCGAGUAGUGUUGGGUCGCGA
 AAGGCCUUGUGGUACUGCCUGAUAGGGUGCUUGCGAGUGCCCCGGGAGGUCU
 CGUAGACCGUGCAU3' (SEQ ID NO: 18)

10 Initial Specific Target Motifs:

- (1) Subdomain IIIc within HCV IRES - nts 213 - 226
 5'AUUUGGGCGUGCCC3' (SEQ ID NO: 19)
- (2) Subdomain IIIId within HCV IRES - nts 241-267
 5'GCCGAGUAGUGUUGGGUCGCGAAAGGC3' (SEQ ID NO: 20)

15

5.8. Ribonuclease P RNA ("RNaseP")

GenBank Accession #s

X15624 Homo sapiens RNaseP H1 RNA:

1 atggcgaggag ggaagtcac cagtgggcc acgagctgag tgcgtctgt cactccactc
 20 61 ccatgtccct tgggaagtc tgagactagg gccagaggcg gccctaacag ggcctctcct
 121 gagcttcagg gaggtgagtt ccagagaaac ggggctccgc gcgaggtcag actgggcagg
 181 agatgccgtg gacccgcccc ttcggggagg gggccggcgg atgcctctt tgcggagct
 241 tggacagac tcacggccag cgaagtgagt tcaatggctg aggtgaggtta ccccgaggg
 301 gacctcataa cccaattcag accactctcc tcggccatt (SEQ ID NO: 21)

25

U64885 Staphylococcus aureus RNaseP (rmB) RNA:

1 gaggaaagtc cgggctcaca cagtcgaga tgattgagt gtcgtgctt gatgaacaa
 61 taatcaagg cattaatitg acggcaatga aataccttaa gtcttcgat atgtagatag
 121 taatttgaaa gtgccacagt gacgtagctt ttatagaat ataaaaggte gaacgcggta
 181 aacccctega gtgagcaatc caaatttgg aggagcactt gttaacgga attcaacgta
 241 taaacgagac acacttcgcg aatgaagtg gtgtagacag atggttatca cctgagttac
 301 agtgtgacta gtgcacgtga tgagtagcat ggaacagaac gcggcttat (SEQ ID NO: 22)

M17569 Escherichia coli RNA component (M1 RNA) of ribonuclease P (mpB)
 gene:

35 1 gaagctgacc agacagtcgc cgcttcgtcg tcgtctctt cgggggagac gggcggaggg

61 gaggaagtc cgggctccat agggcagggt gccagtaac gcctgggggg gaaaccacg
 121 accagtgcaa cagagagcaa accgccgatg ccccgcgcaa ggggagcag gtaagggtga
 181 aagggtgctg taagagcgca ccgcgcggtt ggtaacagtc cgtggcacgg taaactccac
 241 ccggagcaag gccaaatagg gggtcataag gtacgcccc tactgaacc gggtaggctg
 301 ctgagccag tgagcgattg ctggcctaga tgaatgactg tccacgacag aaccgcgctt
 361 atcggtcagt ttacact (SEQ ID NO: 23)

Z70692 Mycobacterium tuberculosis RNaseP (mpB) RNA:
 1 ccaccggta cgtacttgc gaccatggcc ccacaatagg gccggggaga ccgcgcgtca
 61 gtggggcg gcacggtcag taactgtctg gcaacacggg gttgactgac gggcaatac
 121 gggtccatag cgtcgccgc ggatacagta aaggagcatt ctgtgacgga aaagacgcc
 181 gacgacgtct tcaacttgc caaggacgag aagtgcaat atgtcgactg ccggttctgt
 241 gacctgctg gcatcatgca gcaattcacg attccggctt cggctttga caagagcgtg
 301 ttgacgacg gcttggccti tgacggctcg tgatctcgg ggttcacgt gatccaagaa
 361 tccgacatgt tgtttctcc cgtacccgag acggcgcgca tcgaccggt ccgcgcggcc
 421 aagacgtga atataactt ctttgtcac gaccgttca cctggagcc gtactccgc
 481 gcccgcgca acatcgccc caaggccgag aactacctga tcagcactgg catcgccgac
 541 accgcatact tggcgccga ggcggagtgc tacatttgc attcgtgag ctctgactcg
 601 cgcgccaacg gctcctcta cggaggtgag gccatctcgg ggtggtgaa cacgcgcgcg
 661 gcgaccgagg ccgacggcag tcccaaccgg ggtacaagg tcgcgccaaa gggcgggfat
 721 ttccagtgg ccccaacga ccaatagtc gacctgcgcg acaagatgt gaccaactg
 781 atcaactccg gttcatcct ggagaaaggc caccacgagg tggcgacggc cggacagggc
 841 gagatcaact accagtcaa ttgcgtctg cagcgccgcg acgacatgca gttgtacaag
 901 tacatcatca agaaccgc ctggcagaac ggcaaacgg tcactgtcat gcccaagccg
 961 ctgtcggcg acaacgggtc cggcatgca tgtcatcagt cgtctggaa gcagcggggc
 1021 ccgtgatgt acgacgagac gggfllatgc ggtctctcgg acagggccc tcatatcat
 1081 ggcgccctgt tacaccacgc gccgtcgtg ctggccttca ccaaccgac ggtgaactc
 1141 taaacggcg tggttccgg ttacgagggc ccgatcaacc tggctatag ccagcgcaac
 1201 cggtcggcat cgtgtcgcat cccgatcaac ggacgaacc cgaaggccaa gcggtctgag
 1261 ttccgaagcc ccgactcgtc gggcaaccg tatctgggt tctcgccat gctgatggca
 1321 ggcttgagc gtatcaagaa caagatcgag ccgacggcgc cgtcgacaa ggaatctac
 1381 gactcgccg cggaagaggc cgcgagtatc ccgacgactc cgaccagct gtcagatgtg
 1441 atcgaccgtc tcgagggcga ccacgaatac ctcaccgaag gaggggtgtt cacaacgac
 1501 ctgatcgaga cgttgatcag ttcaaacgac gaaaacgaga tcgagccggt caacatcgg
 1561 ccgatccct acgaattcgc gctgtactac gactgttaag gactctcgc agtccgggtg

1621 tagaggggagc ggctgtctgt tggcaggcgc ggcgtcgagg ttcttcgatg ggtgacggtg
 1681 gccggcaacg gcgcgccgac caccgtcgcg aagagcccg ttaagaacgt tcaaggacgt
 1741 ttacgccggg tgccacaacc cgttggcaa tcattcccc accgccgagc ggggtgtct
 5 1801 tcacatgcgc cgaactcaa gccacgtcgt gccccaggcg tgtctcgcg gccggttcag
 1861 gttaatgtc ggggattcgt cgtcggggcg ggcgtccacg ctgaccaacg gggcagtcga
 1921 ctcccgaaca ctttcgcac taccgcttt gccgcgcg tcacccgtag gtagttgtcc
 1981 aggaattccc caccgtcgtc gtttcgccag ccggccgcga ccgcgaccgc attagactgg
 2041 cggccgggtc ccggcagctg gtcggtggcg ttccgcgca ccaacaccg cgcgttgcgg
 10 2101 gccgggttgg cgttcagcca ggcttcaggc agcagctcca cgtcggctcg gggaaacaga
 2161 tcggcgcccg cgtgacatc cagggtttgc agcgtcagg tttgtgcag gccgggaacc
 2221 ttgtgcgcgt gctgtatgtg cagcaactcg acggctcatt cgtatgcggc cagtcgcccg
 2281 cgggccagtt tggttgtgt gttgggtcgc gcaccgcgcg gcaaccgtc ggaactcgata
 2341 cgggcttga tgcggcgaat ctgcgcacc gagtacggc acacaccgtc gggcggatac
 15 2401 cggctttgt cgacatccg taggaatcgc tgaccaact cggcatcgcc ggcacccgcg
 2461 ttgtcgcgta cgaggcccg gatctccat ggctgtgcc acgtctcgta gtatcgcgcg
 2521 taggacccca ggggtcggac cagcggaccg ttgcggccct cgggtcgcga attggcgtcg
 2581 agctccagcg cgggatcgac gctgggtgtc ccacgacgc cccgaacccg ctgcgcgatac
 2641 gatgtcgacc atttaccgc ccgtgcacg tcgacgccg tggccggctc acagacgaac
 20 2701 atcacgtcgg catccgacc gtaccccaac tcggcaccac ccagccgacc catgcccagt
 2761 accgcgatgg ccgccggggc gcgacgtcg tcgggaagge tggccggat catgacgtcc
 2821 agcgcggcct gcagcaccgc caccacacc gacgtcaacg ccggcacaac ctcggtgacc
 2881 tcgagcaggc cgagcaggtc cgcggaaccg atgcgggcca gctctcgacg acgcagcgtg
 2941 cgcgcgccgg cgtggcccg ctccgggtcg ggttagcggc tcgccgagcg gatcagcgcc
 25 3001 cgagccacgg cggcgggctc ggtctcgagc agcttcgggc ccgacggccc gtcctcgta
 3061 ttctggtatg ccgcggcgcg gcgcataac agatccggca catacgcca ggtaccaag
 3121 acatgcatga gccgtttggc caccgggggc ttgtcccgca cgttggccag gtaccagett
 3181 tcggtggcca gcgcctcact gagccggcgg tagggcagca gtccggcttc gggatcgggg
 3241 gcatacgaca tcagtcag cagcctgggc agcagcaccg actgcaccg tcgcggccgg
 30 3301 ccgttttgat tgaccaacgc cgacatgtgt ttaacgcgg tctgcgttc ctcgtagccc
 3361 agcggcgcca gccggcgccc cgcggccttc aacgtatgc cgtggcgcat ctcaacccc
 3421 gtcgggccga tcgattccag cagcgggtga tagaagagt tgggtgttaa cttcgacacc
 3481 cgcacgttct gttcttgag ttctctccgc agcaccgccg ccgcatcgtt ttggccatcg
 3541 ggccggatgt gggccgcgcg ccgcagccag cgcactgct cctcgtcttc gggatcgguu
 35 3601 agcagttggg tgcgttgag ccgctgaac tgcagtcgt gctcgagcag cctgaggaac
 3661 tcatacgacg cggtcattgt ccgcgcgtcc tcacgccga tglagcgcc ttgcccaac

3721 gccgccaatg cgtccaccgt ggacgccacc cgtaacgact cgtcgtctacg ggcatgaacc
 3781 agctgcagta cctgtacggc gaactccacg tcgcgcaatc cgcgcgtgcc gagtttgagc
 3841 tcgcggccgc ggacatcggc gggcaccacg tctccaccc gccgcgcgat ggcttgacc
 5 3901 tcgaccacaa agtcttcgcg ctgcaggct gccacacca tcggcatcaa ggctgtcagg
 3961 taacgctcgc caagtccgc gtgcgcaacg actggccgtg ctctagcaa cgcctgaac
 4021 tccagggtct tggccacgcg ctggtagtag gcgatgtgcg actcgagcgt acggaccacg
 4081 tcccctgtgc gccctccgg acgcaggggcg gcgtccacct cgaanaaggc cgcgagggcc
 4141 acccgcatca tctcgtcggc cagcgcgcgc ttgcgcgggt cggagcgctc ggcaacgaat
 10 4201 atgacatcga cgtcgtcgcg gtatgtcagt tcgcgcgcac cgcacttgcc catcgcgatg
 4261 accgcaggcg gcggtgcggg gtgctcgccg cacacgctcg cctcgccac gcgcagcgcc
 4321 gccgcagagc cggcgtccgc ggcgtccgc aggcgtgcgg ccaccacggg gaatggcagc
 4381 accggttcgt cctcgacct gcgcggcagg tcgagagcgg ccagcattag caogtagtcg
 4441 cggctactggg ttgcgaatcg gtgcacgagc gagcccgga tacctccga ttctcgcagc
 15 4501 cactcgacga acgaccgctg cagctggta tggagcgga gtgtgacct gccccgacgc
 4561 aatttcagg actgcgcatg ggcgaccagg tgaatgcga acgcagcgga cgaagccagc
 4621 accgagaaca gccgcccgcg cagactgcgt tcgcgcagca gagccgcgtt gactcgttc
 4681 catccggtgt ctggatttc cgacagccgg atcaaggcgc gcagcgcggc atcgcgctcc
 4741 ggagcgcgtg acagcgacca cagcagctgc acgtgcgctt gatctcgtg ccgatccac
 20 4801 cccagctgag ccagacgctc accagcaggg gggtaacta atccgagccg gccaacgctg
 4861 ggcaacttcg gccgctgcgt ggcgagttg gtcacgacca cgacggtagc gcaaagcgcg
 4921 tcggcgtcgg atcaaccggt agatcgggc tacagcgaca ggtaggtgcg cagctcgtat
 4981 ggcgctgacgt ggcgtcggtg gtgcgccac tccgtgcgtt tgttcgcaa gaaaaagta
 5041 aaacgtgct ccccaaggc ctccgcgacg agttcgagg cctccatggc gcgcagcgca
 25 5101 ctatccaac tggagcgcaa ttctgggtac ccatcgctc ggctgtcct gggtgtgagg
 5161 tcccatagct tctctcggc ctgcgggcc agcacgtaac cctctctac accccgcaat
 5221 cccgcggcca gcagcagcg gaattgcaga tagggattgc acgcgaatc agggctgcgt
 5281 actcgacc ccgcgacga ggtctgtgc ggcgtgata tcggcaccgc cactagggcg
 5341 gatcggttg cggcccccga cgacgcggcc gtggcgctt cgcgcctct caccagccgc
 30 5401 ttgtaagagt tgaccactg atttgtagc gcgctgatct cgaagcgtg ctccagatc
 5461 ccggtgata acgattacc cactccgac agctgcagcg gatcatcgc gctgtgtaac
 5521 gcgtgacat caccctcgaa cagctcatg tgggtgtgca tcgccgagc cgggtgtcgg
 5581 ccgaatggct tgggatgaa cgacgccgg gcgccctct ccagcgcgac ttctttgatg
 5641 acgtagcgga aggtcatcac gttgtcagc atcgacagag cgtcgccaaa ccgcaggtcg
 35 5701 atctcctgct ggccgggtgc gccttcgtga tggctgaact ccaccgagat gccatgaat
 5761 tcaggggcat cgatcgctg gcggcgaaag tcaagcggg agtcgtgac cgtctgtcg

5821 aaatagccgg cgtgtcgcg cgggacggc accgaccgt cctcgggtcc gggcgtgagc
 5881 aggaagaact cgatttcggg atgcacgtag caggagaagc cgagttgcc ggcctctgtc
 5941 agctgccgcc gcaacacgtg ccgcgggtcc gccacgacg cgaagccgtc cggcatgggtg
 5 6001 atgtcgcaaa acatccgcgc tgaagtgtgg tggccggaac tgggtggcca ggccagcacc
 6061 tgggaagtgc acgggtccgg gtgcgcacc gtaicggatt ccgagaccgc cgaagaccc
 6121 tcgatacagg atccgtcgaa gccgatgcct tctcgaagg cgcctcgag ttgcgtggg
 6181 gcgatggcga ccgacttgag gaaaccgagc acgtctgtga accacagccg gacgaagccg
 6241 atgtcgcgtt ctccagcgt acgaagaacg aatctctct gtccgtccat acctgaaca
 10 6301 gtatgcactg tctgttaaaa ccgtgttacc gatcccgcc cagaagcgtt cggggccggc
 6361 ccgaagggg agtcgcgggt gatttcaggc cgcgcaccgc agactcgtc gcggcaagg
 6421 ccgctcgaga aaatagtga taccgcgaga gtccacacac tgggtgccat cgaacaccgc
 6481 agtgtgttgg gtgccgtcga aggtgatcag cgggtgcgcc agctggccgg ccaggtctac
 6541 ccggactga tacggagtgg ccgggtcgtg ggtgtgtgac accacgacga cttgtccagc
 15 6601 ccggccggcc gccgcgggggt gcggcgtcga cgttcggcc accggccaca gcgcgcacag
 6661 atcgcggggg gcggatccgg tgaactgcc gtagctaagg aacggggcga cgtgacgat
 6721 ccgttgtcgc cgggccacc aggcgcgtgg atcgcggcgt gtggcgcat cgacgcaccg
 6781 gaccgcgttg aacgcgtcct ggtcgtgtct gtagtgcgc tctgcatccc gccgctata
 6841 gtctcggcca agcaccagca agtcgcgggc gtccgtccgc gcctgcagcc ccagcagacc
 20 6901 actggtcagg tacticcage gctgaggcgt gtacagcgcg ttgattgtgc ccgtcgtcgc
 6961 gtccgctgag ctacggccac gtgatccga cgtcttacc ggtctctga ccagcgggtc
 7021 aaccaggcgc tggtagcgtt tgaccactg ggcggagtcg gtgccagag ggccagccgg
 7081 cgagcggcgc cagtcggcgg cgtagtcatt gaaagcgtc tgaatccc ccaattggct
 7141 gatgcttcc tcgattgggc taacggctgg atcgatagc ccgtcgagga ccatgcccg
 25 7201 cacatgagta ccgaaccgtt ccagtaagc ggtgcacca tcggtgcgtt agctgtatcc
 7261 gagtagttg atctgatcgt caactaacgc ttggcgaacc atgtccatgt ccgfgcgac
 7321 ggacgcgcta ccgatatgg ccaagaagct gaagcccatc cgtcaaacac agtctcgggc
 7381 caactcggc tagacctgtt cgacgtgggt gacaccggcc ggactgtagt cggccatcgg
 7441 atcgcggcc tacgcgtcga actcggcgtc gttgcgacac cgcaacgcag gggtcgagt
 30 7501 gcgaaccctc ctccgggtga agcccaccag gtcgaagtgg cggagaatgt ccgtgtcgcg
 7561 gatccgggtt gccatagcgg gcacatgtgc gaccgccgac gccccgggtc cccaaggatt
 7621 gacagcagt gctcgaatc gctgtccgt cgcggggacg cggatccgc ccaactcgc
 7681 ttgtgtcca ccgggttgtt cgtatcgac ggggacggac accgtcgcgc agcgtcgagt
 7741 gcgaatttc ctgggtcgg cgatgaact gcggcagctg ttcaactct gtcggcgcg
 35 7801 cagcaccggc gcacccgggg ttggccggc gccgggttct tcagtgcgc cggccaacgg
 7861 gggcgtcgt aggggcagtc gcgcgagcag caaccgaag gacagcagc ccgagctcaa

7921 cggctctcgg cgccacatgg ccgccatcgt ctaccggcg aatacctgtg acggcgcgaa
 7981 atgatcacac ctctgtttt tcgcccctg agcaactggc gccgctgggc ggctgctgac
 8041 cggcgattaa atacgcctgc acgtactcgt caatgcagct gtcgccctgg aataccaccg
 5 8101 tgtgctgggt tcgctgaag gtcagcaacg aaccgcgaag ctggttcgac aggtcgaccc
 8161 cggccttgta cggcgtcgcc gggtcatggg tgggtgatac caccaccgtc ggctactaggc
 8221 cggcgccgga gacggcatgg ggctgacttg tgggtggcac cggccagaac ggcaggtgac
 8281 ccagcggcgc atcacgggtg aactccctg agctcatgaa cgggtcgatc tcccgggcgc
 8341 ggcggtcttc gtcgatgacc ttgtcgcgat cggtaaccgg gggtgatcgc acgcaattga
 10 8401 tcgccaccgg cgcgtcacgg gaattgtgt agcggccgtg cgagtccga cgcattgata
 8461 tgtcggccag agccagcagg gttgtccgc gattgtgac cagctccgac agcccgtcgc
 8521 tcaagtgttg ccacagatc ggtgtagata gcgccataat ggtgccacg atggcgtcgc
 8581 tataactcag cccgcgcgga tcttcgtgc gcgccgctc gctgacctc ggtgttcgc
 8641 ggtcgcacaa cggatcgacc aggtctgtgt agacctgcac ggtttggcc gggtcgggcg
 15 8701 ccagcgggca gcccgcttc ttggcgcagt cggcggcata gttgtgaac ggcctctgga
 8761 acgctctggc ctggcgcagc tccgctcga tgggacggc attggggtcg acggcaccgt
 8821 cgagaatcat tgcccgcacc cgtgcgggaa attcctcggc atacgcggag ccgatccggg
 8881 tggctacga gtagccagg taggtcagct tgtcgtgcc caacgccgcg cgaatggcat
 8941 ccaggtctct ggcgagttg accgtccga catgggccag aaagtcttg ccatcttgt
 9001 ccacacagcg accgacgaat tgcctgtct cgttctgat gtcgccaca cctcccgcc
 20 9061 tgtagtcaac ctgcgctcg gcccgagcc ggtcgtgtc ggcatcgag ttgcaccaga
 9121 tcgccggccg ggacgagcc accccgggg gylcgaacc aaccaggtc aaactttgt
 9181 gcaccgctt cggcaatgic tggaaagcg caagggcgc ctcgataccg gattcgccgg
 9241 gtccaccggg atttatgacc agcgaaccga tctgtctcc cgtcgccgga aagcgaatca
 25 9301 gcgccagcg cgccacgta ccatcggggc ggtcgtatg gaccggtaca gcgagcttg
 9361 cgcataacgc gccgcgggg atctttact gcgggttga cgaccggac ggtgtccact
 9421 ccaccggctg gccacgttc ggctccgcca tacgagcgcg tcccgcacc acgcggtatc
 9481 agcccacaa aaccaagcc acggcgccga gcgcggccca galcaacagc atgcgcgga
 9541 tctgtcgcg gcgagacagc ctcatgcca caatgtgac agagcagacc cgagatctg
 30 9601 ccagcgggcc accgtcggc gactaaccgg ccgctgccag cagtctgac atcgccgatg
 9661 gcgaactcgt cggccatccc ccatactgc ggtaacagat ccgggcaaga caccgacccg
 9721 tcgaccggat ccggcacggg cgcgtcgcc tcggcggtgc acaactcgca catcaggttg
 9781 gcgtggcac cccgtccacg ccgcatggt gcacctggc catgcocga gggcgatcc
 9841 cgtgcgtgc caccctcg acgaacctc ctcccaggg ggtcgccgc agcgacgca
 35 9901 tgtggcgca gatcccgag agttcgccc gcccgccgg cgacggcaac ccgatgccgt
 9961 gcaagtgcg atcgatgta ggttaaggt tcagcgaact gctggaagc ttttccgaa

10021 accgcggcct cgccttgatc tggagtcaga acgcgtcacg cagccggfca aagcgctaac
 10081 ccattgctga gcaaacatgc atgggctgag tggacgttfc cagacacagc aactggcgtc
 10141 caggccactg agccgctgca tggcgcgatg tatgccgatg ggggccccgg gcgcgtctga
 5 10201 ggggaagaag tggcagactg tcaaggctcg acgaaccogg gggacctaac gggccacgag
 10261 gatgcacccg accaccatta gggacagtga tgtctgagca gactatctat ggggccaata
 10321 ccccgaggag ctccgggccc cggaccaaga tcgcaccca ccacctacag agatgnaagg
 10381 ccgacggcca caagtgggcc atgctgacgg cctacgacta tlcgacggcc cggatcttgc
 10441 acgaggccgg catccgggtg ctgctgctgc gtgattcggc ggccaacgtc gtgtacggct
 10 10501 acgacaccac cgtgccgac tcattcgacg agctgaccc gctggtccgt gggctgtgtg
 10561 ggggtgcccc gcacgcactg gtctgctccg acctgccgt cggcagctac gaggcggggc
 10621 ccaccgccgc gtggccgccg gccaccogggt tctcaaggga cggcggcgca catgctgctca
 10681 agctcgaggc cgggtgagcgg gtggccgagc aaatgcctg tctgaccgcg gcgggcaccc
 10741 cgggtgatgc acacatcgcc ttaccccgcc aaagcgtcaa cacctgggc ggtctccggg
 15 10801 tgcaggccgc cggcgacgcc gccgaacaaa ccatgccga cgcgatgcc gtcgccgaag
 10861 ccggaaggct tccgtctgtg atggagatgg tggccggcca gtggccacc cagatcaccg
 10921 gcaagcttac cattccgacg gtgcggatgc gcgtgggcc caactgcgac ggcacggtcc
 10981 tggatggcca ggacatggcc ggggtacgag gcgccaaagc cggccgttc gtaaacggt
 11041 atgccgatgt cgggtgtgaa ctacgccgtg ctgcaatgca ataccgcaa gagggtggcg
 20 11101 ggggggtatt cccgctgac gaacacagtt tctgacaag ccgaatcag ccgatcgccg
 11161 ggcattgcgg tggcgccctg gatgccgtg acggcggatt cggcgccgc gg acgcgccagc
 11221 gggaccatc ggcgtcgcgt tcggcggtg agccgggggt gagccagac attcgatgtg
 11281 cccaacacca tcgccacag ccgaattgat gtggcactct atgcatgct atcccgcac
 11341 aaccaccacc gcggcgacgc atcatgaccg gaggcggaaga tggcagtaga ggcgccgaga
 25 11401 ccagcgcgcc atctgaggtg cgaagcgcaag ttcgacgtga tggatcgac ggtgtcgccg
 11461 tctgtcagg gcacgcgcc ggtgtgtcgc gtgcgacgt cggcgacca cgactcgac
 11521 ggggtgtact togacacacc gtgcgacgac ctggcgcgca accagatcac ctggcgccg
 11581 cgcaccggcg gcggcgacgc cggctggcat ctgaagctgc cggcgccacc cgaagaagcg
 11641 accagatgc gaggaccgt gtccgcatc ggcgacgtg tggcgccga gtgttgat
 30 11701 gtgtgtctgg cgatcgctcg cgaccagccg gttagccgg tgcggcgat cagcactcac
 11761 cgcgaagacc agatctctga cggcgccggg ggcgacgagc tggcggaatt ctgcaacgac
 11821 gacgtcacgc catgctggc cggggcattc cagcgccgtg gtgcagcgga caacggccct
 11881 gccgaacagc agtggcgca atgggaactg gaactgtga ccacgcatgg gaccggcat
 11941 acaagctac tggaccggct agccaaccgg ctgctcgatg ccggtgcgc acctgcggc
 35 12001 cagcgctcca aactggcgcg ggtgtcgggt ggcacctc ccggtgagc gcccaaggc
 12061 ccgcagccgc cggcgcatcc agtacaccgc gcgtgtccg agcaagtga gcactgctg

12121 ctgtggggtc gggcgctgcg gggcgacgcc taigacgccg tgcaccagat gcgagtgcg
 12181 accgcgaaga tccgcagctt gctgacgat tccaggagt cgtttggcct gaaggaaagt
 12241 gcgtgggtca tcatgaact gcgtgagctg gccgatgtcc tgggcgtlga ccggagacgc
 5 12301 gaggtactcg gtgaccgcta ccagcgcgaa ctggacgcgc tggcgccgga gctgtglacgc
 12361 gggcggtgtgc gcgagcgctt ggtagacggg gcggcgccgc gataccagac cgggtgctgg
 12421 cgtactacta tcgcatgtcg gtcgacgcgg tacttcctgc tgcctgacgc tctagacgcg
 12481 ctgtgtgtcc aacgcgcca tgcacttct ggggaggaaat cggcaccggt aacctcgtat
 12541 gcggcgctacc ggcgagtcgg caaagccgca aaagccgcaa agaccgccgg cgacagcggg
 10 12601 ggcgaccacc accgcgacga ggcattgcac ctgatccgca agcgcgcgaa ggcattacgc
 12661 tacaccgcgg cggtacttgg ggcggacaat gtgtcacaag aagccaaggt catccagacg
 12721 ttgctaggcg atcatcaaga cagcgtgtgc agccgggaac atctatcca cgagcccata
 12781 gccgcgaaca ccgccggcga ggacaccttc acctacggtc tgcctacca acaggaagcc
 12841 gaattggccg agcgtgccc ggcgagcgtt gaagccgcgc tgcgcaaat cgacaaggcg
 15 12901 gtccgcaaa gacgggattg agcccgcag gggcgacga gttggcgtt aagccggatt
 12961 ctgttccgcg ccgccacagc caagctaac gcggcacggc ggcgaccatc catctgaca
 13021 caccgttacc ggggtgctcg agcggcttac ccgacggctc gggcgagcaa cctcgaagcg
 13081 cctgcgcgcg cgcacttctg gtgcgcctt ctggccttg cttgggtgg ggtttgccta
 13141 gccacccggg tcaccggaa tgctgtgtcg ctcttaccgc accgtttcac cttgccacc
 20 13201 acgaggatgg cgtctgttt tctgtggcac ttcccgcga gtcacctgg attgcggtta
 13261 gcaatcaccc tgcctgtga agtccggaat ttcttgcagt cagcgtgaa cctcgtgaat
 13321 ccacacaagc cctacgcgag ccgcggccgc ccagccaact catccgcgac gaccacgcta
 13381 ccccgtctgg cgtgtgcgc gccagtgtga ccgctggacg acacggctag tggacagcc
 13441 gatccggcgg gcagtcctta tctgtgactg gtgacacgtt gggacaaacg cgtgactcc
 25 13501 ggcgacttgg acgccatcgc tgcgaggtc agcgagtacg gtggcgcaat gctacctggy
 13561 ctgatcaccc ccggcgagcg cgcccgctg cgaagctgt accccgacga cggcctgttt
 13621 cgctgcagcg tcatatggc atccaagcgg tacggccgc ggcagtatcg atattccat
 13681 gccccttate ccgagtgcac gagcgtctca agcaggcgct gtatoccaa ctcgtccga
 13741 tagcgcgcaa ctgttgggcc aaactgggcc gggaggcgcc ctggccagac agccttgatg
 30 13801 actgtttggc gagctgtcat gccgccggcc aaaccgacac cacagcgctg atgttgaagt
 13861 acggcaccaa cgactggaac gccctacacc aggaletcta cggcgagtfg gtgtttccgc
 13921 tgcagggtgt gatcaacctg agcgtaccgg aaaccgacta caccggcgcg gagtctctgc
 13981 ttgtcgaaca ccggcctcgc gcccaatccc ggggtaccgc aatgcaact ccgcaggac
 14041 atgttatgt gtgcacgacc cgtgatcggc cgtgtcgguac tagccgtggc tggtcgcat
 35 14101 ctccagtgcg ccatgggctt tgcatactt gttccggcga acgctatgcc atggggctga
 14161 tctttcaga cgcagcctga ttgcacgcca tctatagata gcctgtctga ttaccaatc

14221 gcaaccgacga tgccccatcg gcgtagaact cggcgatgct cagcgaatgc agatcaagat
 14281 gcaaccgata taggagcccg gaccggcat ccaacgcag cgcacaacat attttgatcg
 14341 cggtagatcg tgaccacc agcaccgtcg cgcttcgta gccacgatg atcgcgatcac
 5 14401 gtccccccg aaccggccg agcacgtcgt cgaagcttc cccaccggg ggcgtgatgc
 14461 tgggtgctcg cagccagcga cgggtcagct cgggatcgcg ttctcgccgc tcgcgaacg
 14521 tcagcccttc ccaggcccg aagtcggtct gcaccaggtc gtatcgacg accacgtcca
 14581 gggccaggcg tcggcgccg gtcaccggg tctcgtaac ccgctgtagc ggcgaggaga
 14641 ccaccgcagc gatcccgcg cgcgcgcga gataccggc cgcgcacca acctggccgc
 10 14701 accccacctc gttaacccc ggggttccgc gccccgaata gcggcgctgc tcgcagact
 14761 cgtctgccc gtggcgcaac aaaagtagtc ggggtgggtg accgcggcg cgcgacgc
 14821 cgggagatgt cggtagtcg gtcgcaacga ttggcgagg atccgatcc gccgcagccg
 14881 attgcgcgc ggcgtccatc gcgtattgg ccaaccggc tgcatacgtg ttccgggcac
 14941 gcggaacca ctctagtgg atccgcgaa actgggacgc caacgcctga gcctggacat
 15 15001 agagcttcag cagatccggg tcttgacct tcaccgcgc ggacatcgc tccaccacca
 15061 gcttgagtc catcagcac cggcgctcgg tggcacctag ttacggcg cgtccaacaa
 15121 cggctatcag ccgcgcgtat tcggcgactg tgttcgtcg ccggccgaic gcctgcttgg
 15181 actcggccag cacggtggag tgatcgccg tccaccac cgcgcctgat ccggccggtc
 15241 cgggattgcc ccgcgatcc ccgtcggctt cgaagacaac ttactctc caaatcttc
 20 15301 gagccgaac aagatcgctc cgcattccg gcagcgacc acttactct cggcgccgc
 15361 cgagatctg gccagctcg cgcggccgat ctgatccgg caggcaccac atcgatgac
 15421 ttgcaaccg ccggccctcg gccgcctcc gcccgctgt cttcgtaga gcccgcaag
 15481 ctgcggatca agtgcgcg tcagcatgc gcgttcgat gaatgttgt gccggcttg
 15541 gtcgatttc gcaagtgcct cgtccaaagc ctgctggcg cgcggcaggt cggcccgaa
 25 15601 cgcttgagc gcccgcgact cggcggtctg ttgacctgc agctctcgc ggcgttcag
 15661 caccicagc agggcatctt ccaactggc ttgacggct tgcaagctgt cgaagctctg
 15721 ctgcagatca gccattgct tggctccgt tgcaaccgaa gtgagcaac accggtccc
 15781 gtcgccagc ttacgcacc catcgatc cgactcaaa cgcgcacct ggcgtccaa
 15841 gtccicgcc gcgattgca gggccgcat cctgtcgtg cggcgctgt gctcgctc
 30 15901 cactgctcg taagccgcc gtcgcccag atgggtagcc cgaatgcga tccgggtcag
 15961 ctacgatcc agcttcgca attcagtag cgaccgttc ttgcccact cggcttcat
 16021 gcctgatc tcaccgttc gtatcgagg ttccagggt cgtgcagat ggtgcacaca
 16081 cgcaccgca cgcagcgcc gaattgagc cgcaactt cggcgctg gccgcaccac
 16141 ggggaattgc ttgccaatg cgcagctcg atcaggcca ctgcaagc tcggcaatgc
 35 16201 tctcggtcg gatgatcgc cagatcgcc gtaacgtac ctgacgcgc cgcgcggcc
 16261 acgttgccaa gcaacgagc cccggccgc ccgcagacc cgaccgcga caccgagcg

16321 tcgggatccc cgggcgcgcg cacaccggtc gcagtcggcg gcaacgcggc ctccagacgg
 16381 gaacaaaagg tgcgcagcgg ttccgggttt ggcagtctgc caatccggcc taaccgcgtg
 16441 ccgaccggcg gtgtgtaccg cgcgaagatg tcgaatgccg gctcctctga aggggtgcgcg
 5 16501 gcgcgcacgc ccgccaacac ctggcgcgcg gctcgtgcgg gtgcgcagac ctgcaccggg
 16561 tctctggcca ccgttcgac ggtagccgac ctgcctatgg cggcgcgacgc cccgtcgtgc
 16621 gccaggaact gcccggtacc cgcgacacgc cagctgcagt gcgagtagtc gccgatatgg
 16681 ccggcaccgg cctcaaagac cgtgcctccg accgcctctg agttctcgcg cggcacatag
 16741 atgaccacact tgtcgagatc ggccgctccg ggcaaccggg cgagaacggc gtgcagcggc
 10 16801 agaacaacag cgtgtgccag gcgcgtggac acacccggcg acggcgagtc ggcggttggg
 16861 tgcgcggtaa acaacgagcg accggtcggc atcaggcggg gcaccagcac acccttggc
 16921 gtgttggcgg cgaccgtatc gaccceacgc agtaacaagg ggtgtgtgac caatagcagt
 16981 ccggcctggg gaacctgtgc caccaccgcc ggctgcgcgt ccaccgaac ggtagccgaa
 17041 tccaccacgt cgtcggggtc gccgcacacc agaccaccgc aatcccacga ctgggcaagc
 15 17101 cgcggcgggg aggcctgtgc cagcacgtgc atgacatcgg ccagccgcac actcatcggc
 17161 gtctccacg cttgtccac tcgctgatcg ccgcaccag cacgggccac tccgggcga
 17221 ccgcgcgccg caggtagccg gcgtccaggc cgacgaaggt gtcaccgcgc cgcaccgcaa
 17281 ttctttgct ctgcaatag ttctgtaac cgtcagcalt ggcatgttg aacagtacga
 17341 aaggggcgcg accatcgacc acctcggcac ccaccgact cagtccggcc accatctccg
 20 17401 cgcgcagcgc cgtcaaccgc accgcacggc ctgcggcagc ggccaccgcc cggggggcgc
 17461 agcaagcagc gatggccgtc agttgcaatg ttcccaacgg ccagtgcgct cgtcgcacgg
 17521 tcaaccgagc cagcacgtct ggcgagccga gcgcgtagcc caccgcgaat ccggccnagg
 17581 accacgtttt cgtcaagcta cggagcacca gcacatcggg cagcgagtca tcggccaacg
 17641 attcggctc gccgggaacc caatcagcga acgcctcgtc gaccaccagg atgcgtccg
 25 17701 gccggcgtaa ctgcagcagc tgcctcgga ggtgcagcac caggtgggg ttgctcggt
 17761 taccacgac gacaaggtcg gcgtcgtcag gcacgtgcg ggtgtccagc acgaacggc
 17821 gcttaggac aacatgtgc gccgtgatic cggcagcgct caaggtatg gccgctcgg
 17881 tgaacgcggc cagcagcalt gctcccgca ccggacttag gttgtgcagc aatgcgaatc
 17941 cctccgccgc ccgcagcagc ggagacactt cgtcacgggt totgcatga cgttcagcga
 30 18001 ccgcgtcttg cgcgggtgc acatcgtcgg tgcctggata cggggccagc tccggcagca
 18061 gcgcgcgcag ctgcgggacc aaccattcgg ggggcccggc atggcgagc ttgacggcga
 18121 agtcagcac gccgggcgcg acatcctgat caccgtgta gcgcgcgcg gcaagcgggc
 18181 tagtgtctag actgcacaca gcgtcaaca gtatgggcc ggtgtcggg ccaagaaatc
 18241 agagcaccgc cgcgcgttg tctacgggc gacaaccgcg acatcacagg cagtaacag
 35 18301 ggctgcgcg gtgatgatc tcaggccaag cagctgtgcc tgggcgatga gcacacgtc
 18361 gaatggatgt cgtggtgat ccggaagctc tgcgtgcgc agtgtgtgcg ttgtcaactg

18421 acagcgcgca cgtgccgcag cggcgcatc gatcgggcac gtaagaagcc galgctcgg
 18481 gcgccgggag cttgccgagg cggtagtga tcgcgatc ccaggcactg gcggccgaca
 18541 agagaatgct gttgccgacg tcctgaacaa tcgccgtgt ttggtgacg gcatccgcag
 5 18601 ccaaacgtgg gtgtcgatga ggtagcgctt caccggtgaa agcgtlccag cagtcgtct
 18661 gacaacggag cgtccaaatc gtcgggcacg cggtagacgc catgttcaat gctaacccg
 18721 cgagtctcat gaggatgcag cggcacaagc ttgtctaccg gctcgcgcg cggggcaatc
 18781 tcaacctctg cccgccgtag acgagccgca gcagctcgga caggcgtgtc ttcgctcgt
 18841 gaacgccgac ccgcttcgca ggccgccaga ctttcgcgtc gaccactgc tcacaaact
 10 18901 tcgcgatcat cgcctgatac cacagcgcca accggtagcg gttgtccaa ccgcttcgc
 18961 aacgacaatg ggatcgtgac cgacacgacc gcgagcggga ccaattgcc gctctccca
 19021 cgcgccgccg caggcgccgc atcgtccgcg ggtagaatgc cgcagctggt gatcttcgat
 19081 ctggaccgga cgtgaccga ctgcgcgcg ggaatcgat ccagcttccg aacgcgcgc
 19141 aaccacatcg gtgccccagt acccgaaagg gacctggcca ctcaatcgt cggcccgcc
 15 19201 atgcatgaga cgtcgcgcgc catggggctc ggcaaatcc ccgaggaggc gatcgtagcc
 19261 taccggccgc actacagcgc ccgcggttgg gcgatgaaca gcttgttga cgggatacgg
 19321 ccgctgctgg ccgacctgcg caccgcggtg gtcggctgg ccgtgccac ctcaaaggca
 19381 gagccgaccg caccgcgaat cctgcgcac ttcgaattg agcagcactt cgaggtcatc
 19441 gcggcgccga gcaccgatgg ctgcgaggc agcaaggctg acgtgtctgc ccacgcgctc
 20 19501 gcgcagctgc ggccgtacc cgagcggttg gtgatggtc gcgaccgag ccacgacgtc
 19561 gacggggcgg ccgcgcacgg catcgacacg gtgtgtgtc gctggggcta cgggcgcgc
 19621 gactttatcg acaagacctc caccaccgtc gtgacgcat ccgccagat tgacgagctg
 19681 agggaggcgc taggtgtctg atccgtgca cgtcacatc gttgtacgg gcaacatctg
 19741 ccggtcgcca atggccgaga agatgttcgc ccaacagctt cggcaccgtg gctcgggtga
 25 19801 cgcggtgcga gtgaccagtg cgggcaccgg gaactggcat gttagcagtt gcgcgagca
 19861 gcggcgccgc ggggtgttc gagccaccg ctacctacc gaccaccgg ccgacaaagt
 19921 cggcaccgaa cacctgcggc gagacctgtt ggtggccttg gaccgcaac acgctcggt
 19981 gttcggcag ctgcggctcg aagccgcgcg ggtacggatg ctgcgctcat tgacccacg
 20041 ctggggaacc catgcctcg atgtcgagga tcctactat ggcgatcact ccgacttga
 30 20101 ggaagtcttc gccgtcatcg aatccgccct gcccgccgtg cagactggg tcgacgaacg
 20161 tctcgcggcg aacggaccga gtgtatgcc cgcctagcgt tctgtctgc gcccgcttgg
 20221 ctggcgcttg ccttggtctg ggtcgcgttc acctactgt gctttacgtt gctcgcgcg
 20281 tggcagctgg gcaagaatgc caaaacttca cgagagaacc agcagatcag gtattccctc
 20341 gacaccccg cggttccgtt gaaaacctt ctaccacagc aggatcgtc ggcgccggac
 35 20401 gcgcagtgcc gccgggtgac ggcaaccgga cagtacctc cggagctgca ggtgtcggc
 20461 cgactcgcg tggtagaggg ggaccaggcg tttaggtgt tggcccatc cgtgtcgcg

20521 gggggaccaa cggtcctggt cgaccgtgga tacgtgcggc ccagggtggc ctgcacgta
 20581 ccacgatcc cccgcctgcc ggtgcagacg gtgacctca ccgcgcggct gcgtgactcc
 20641 gaaccgagcg tggcggggcaa agaccattc gtcagagacg gcttcagca ggtgtattcg
 5 20701 atcaataccg gacaggctgc cgcgcgtacc ggaagtccagc tggctgggtc ctatctgcag
 20761 ttgatcgaag accaaccggc cgggctcggc gtgctcggcg ttccgcatct agatcccggg
 20821 ccgttcctgt cctatggcat ccaatggatc tcgttcggca ttctggcacc gatcggcttg
 20881 ggctatttgc cctacggcga galccgggcg cgcgcggggg aaaaagcggg gtgcaccaca
 20941 ccggacaagc caatgacggg cgagcagaaa ctgcgtgacc gctacggccc ccggcggta
 10 21001 accaacaatca cggccaatc cgcagccccc gctgggacca ccgcgcagc caccaggcg
 21061 cggcgagat cggccacctt gggcgaccgg ccgtgcacca aggtgggccc gatctgcaac
 21121 tcatgttgtt accgggtggg ccaaccagc cgcactcaa ggcgccagc aaacgcggcc
 21181 tcgacgacac cggcggtggg gctgggatgg cggcgggcgt cgcgcgcaca gggccgtacc
 21241 gcaccggggg cgcaccacc gaccaccggc gcgcagatca ccaccagcac cgcctcggcc
 15 21301 cgtgcgcaa catagtggc ccagtcacc aatcgtctg cagcccaacc gaatcggaga
 21361 taacgcggcg agcggtagcc gatcatcgag tccagggtgt tgatggcagc atatccagc
 21421 accgcaggca cgcgcctcga agcgcgccac agcagcgcca ccaccgggc gtgcgcggtg
 21481 ttctcgcca ccgactccag cgcggcacgc gtcaggccc ggcgcgccag ctgggcgggg
 21541 tcagccccc acagcgacgg cagcagccgt cgcgcggcct cgacatcgt cgcctccaac
 20 21601 aggtccgata tctggcgcc ggtgcgcgcc agcgaagtc cgcaccgac tggccaggtg
 21661 gccgtcggc tggccggcac gggccaggac ctgcgggta ggcgtgcag tggccggcg
 21721 agcaagccca ccgcgcgac cagcaggccg acgtgtaccg caccggcgac ccggcgctca
 21781 cggtagggtg tctgtccag ctggcgccc gccgcagc acaggccac cggatgacct
 21841 cgttgggggt cgcgaacac gactcggagc aggcagccga tcagcacgcc gacggccctg
 25 21901 gctgcaggc tcatgcaaa cactccggca cgtcgcaca cgtggtctac gtcagctat
 21961 ttatgacct atacggcagc tatccagat gaagcgcca gctaccggg ttggccacct
 22021 gtgaacccc gggcaatgt tgttcgggc agcgaatgc atcatgcagc tggcagtgcc
 22081 ggggtcggg tatggcgtgc tggaaagccc ggtgcagacc ggaacgtct acaagcatcc
 22141 gttcaagcgg gcccgacca ccggcaccta cctgggggtg cgcaccatcg ggaaggatc
 30 22201 gcaccgagcg ctgatccggg gtgccttggg cgtcgcgac cgcaggttc gttcgcaggc
 22261 ctcagcccca gtgtclata acgctctga ccggaagtg cagctgtggg tggcgcggtg
 22321 tctgtaccgc tacttctgg accagcaga gtttctgtac ggccactcg aagatgcac
 22381 cgcgcagccc gctaccaag acgccaacg gttaggagac acgtgcagg tggcgaggg
 22441 gatgtggcc cgggaccggg tcgcttcca cgaagtctgg aagcgtcgc ttgatggct
 35 22501 gcagatcgac gcgcgggtgc gcgagcatct tcgcggggtg gcctcggtag cgtttccccc
 22561 gtggccgttg cgcgcgggtg ccgggocgtt caactgttt gcgacgacg gatcttggc

22621 accggaggttc cgcgcgatga tgcagctgga ggtgtcacag gccccagcagc gtcgcttcga
 22681 gttggttaact tccgtctac ggtagccga ccgctgatt ccgcatcggg ccttgatctt
 22741 cgtttaccag ctttactgt gggacatcgc gtttcgcgc ccacacggcc gccgaatcgt
 5 22801 ctgatagagc ccggccgagt gtgagcctga cagcccagaca ccggcggcgt gttgcgcgtc
 22861 gccaggttca cgtctggcga tctagagccg ccgaaaacct actctgggt tgcctccga
 22921 atcaacgtgc tgatctgctc gaggcgtca cgcatacgg cgcgcacgc atccaccgcg
 22981 gcatacaggt cggccttggc gcccggcagc tggctccgac tcaftggccg caccggcggc
 23041 gctgtctgtc gcggccgctc gtcgcttga aaccaggtc gtcacccac gaccagaca
 10 23101 ctgccatc cggcgcggcc ccgacaacga agcacagcta gccggtgggc ggcgacggga
 23161 tcgaaccgcc gaccgctggt gtgtaaaacc agagctctac cgtgagcga cgcgcccatg
 23221 accgcccgag gctacacgcc ttgcggccaa gcaocaaaaa ccttaggccg taagcggccg
 23281 cagagcgtgc gtcacagcc gctgacgcg aacticacc ggctgcttca tctcggcga
 23341 ccgaatgatc cctgaccgat cgaccacaa ggtgccggcg ttacgatgc cggcctgtc
 15 23401 gttgaagacg ccgtaggcct gactgaccg gccgtgtggc cagaagtccg acaacagcgg
 23461 aaacgtgaat ccgctctgcg tcgccagat cttgtgagt ggtggcgggc ccaccgaat
 23521 cgtctagcgc gcgtgtcgt cgttctcaaa ctgggcagg tgaacacga actggtccag
 23581 ctgcctcgg cagatgcgcg tgaacgcaa cggaagaac accaacagca cgtttctgc
 23641 accccggtag ccgcgcaggg tgacaagctg ctgattctgg tcgcgaacg tgaagttagg
 20 23701 ggcggtggct ccgaggtca gcatcagcc ttgccagcc gcgatttcg cgtaccaat
 23761 ctgtggcgcc tcagttgcc cagatgacc gacgagtcg gcatcagccc agctgtgggc
 23821 gccgcctcgg caatctcggc gggcaataca tggccgggct ggcggtctt ggcgcacc
 23881 acccaaatca caccgtctc ggcgagcggc ccgacgcga ccatcagggt gtcaccaa
 23941 tcgccgtgc catcagcca ccacaacagg acgacatga tgacctcgc ggtgtcttca
 25 24001 tcgagcaact ctcccccga cgtttctg atggccgcgc ggaatgtctc gtcgtgtct
 24061 tcgtccagc ccattcctg gataagtgg tctgttga tggccaatt gcggcgtag
 24121 ttgagggct gatccgcgc gaccacgtg gaactcctt cagtctccg gggocatgtg
 24181 cacaccgtc cgtgggcat tatcgtcga cagcagaac cgttcaccc gccgcctca
 24241 gaagcggcc accgacattg tcaatgcctt tctcttggc tcttgagcc gatcaaccg
 30 24301 ccggttgaat tcgctgtcgc acgcgtgcgc accgatgga ttgacaccg cgcggggccg
 24361 gtcgacatat gcgttagcg catccccag ttgcgagac agcgcggcgc tcagactgcc
 24421 tgaaccgtc gaggcactgt tgttagcgc gtcgatgcc ggaactcgg tcggccgggt
 24481 gttgcggccc tgaftgaacg cggccacgta ggcgttacc ttgtgatgg cgtctgtct
 24541 ggtggccgc agcgcgtcac acgaggtgcg aatgccttg gtcgtcagc atftgtggc
 35 24601 ctgcactcc ccgatgtcgc acgtgcgcg ogaagccgac accgacggc acaccagca
 24661 gcggtaggcc ggtgcgactt tgggtcggg catggccgta cgtcgggtga cagtgtgata

24721 tccgacgac cccatcagca gcagcgcgat gcagccgagc gccaggcgcg ctcgcctggg
 24781 gagctcccc ccgtgcctgc gaggcacggc gcgccatccg atgagcacgg catgtgaggt
 24841 tacctggctc gagcgcgacc gcgctggccg tgggtgtctc cgcatccgca gaaccgagcg
 5 24901 gagtgcggct atccgccgcc gacgccggtg cggcacgata gggggagcac catctaaaca
 24961 gcacgcaagc ggaaagccgc caccctacgg agtagtgcgt tgaccaccga ttctgccgcg
 25021 cagcatctgc cccaaaacte aaacagcgca agcgaaccgc accgagttcg gttgatccgc
 25081 gaggggttgg cgtcgtattt gcccgacatt gatcccagg agacctcgga gtgctggag
 25141 tcccttgaca cgtcgtcgca acgtcgccg ccgtcgccgg ccgctacct gatgttcgg
 10 25201 ctgtagagc gggccggcga gcagcgggtg gccatccgg catgacgtc taccgactat
 25261 gtcaacacca tcccgaccga cgtggagccg tggttcccc gcgacgaaga cgtgaacgt
 25321 cgttatcgag cgtggatcag atggaatcg gccatcatgg tgcaccgtgc gcaacgaccg
 25381 ggtgtggggc tgggtggcca tatctgacc tacgcgtct ccgcggcgct ctatgaggtc
 25441 ggtttcaace actctctcg cggcaagtcg caccggcgcg gcggcgatca ggtgttcate
 15 25501 caggggccacg ctccccggg aatctacgcg cgcgccttcc tcgaaggcg gttgaccgcc
 25561 gagaactcg acgattccg ccaggaacac agccatctc gcggcggtt gccctctat
 25621 ccgcacccgc ggctcatgcc cgactctcgg gaattccca ccgtctgat ggggttggc
 25681 ccgctcaacg ccatctacca ggcacgggtc aaccactate tgcatgaccg cggatataaa
 25741 gaacctccg atcaaacgt gtggtgttt ttggcgacg gcgagatgga cgaaccggag
 20 25801 agccgtgggc tggccacgt cggcgcgctg gaaggcttg acaactgac ctctgtgat
 25861 aactgcaate tgcagcgact cgacggcccg gtgcggcgca acggcaagat catccaggag
 25921 ctgagtcgt tcttccgcg tgcggctgg aacgtcatca aggtggtgtg gggcgcgaa
 25981 tgggatgcc tctcgcagc cgaccgcgac ggtgcgttg tgaattaat gaatacaaca
 26041 ccgatggcg attaccagac ctataaggcc aacgacggcg gctacgtcgc tgaccacttc
 25 26101 ttggccgcg accacgcac caaggcgctg gtggagaaca tgagcgacca ggtatctgg
 26161 aacctcaaac gggcgggcca cgattaccg aaggtttacg ccgcctaccg cgcgccctc
 26221 gaccacaagg gacagccgac ggtgatctcg gccaaacca tcaaggcta cgcctgggg
 26281 aagcatttcg aaggacgcaa tgcacccac cagatgaaa aactgacct ggaagacct
 26341 aaggagtctc gtgacagca cgggattcgg gtacggagc ccagcttga agagaatccg
 30 26401 tacctgccg cctactacca ccccgccctc aacggccggg agattcgta catgtcgac
 26461 cggcgccgg ccctggggg cttgttccc gagcgaggc ccaagtcгаа agcgtgacc
 26521 ctgcgggctc ggcacatcta cgcggcgctg aaaaagggt cttggcacca gtaggtggcc
 26581 accacattg cgacggtcgc cagcttcaaa gaagtgttc gcgacaagca gatcggggcg
 26641 cggatagtc cgatattcc cgacgagcc cgacacctg gtagtgactc ctgttccc
 35 26701 tctetaaaga tctataaccg caatggccag ctgtataccg cggttgacgc cgacctgatg
 26761 ctgcctaca aggagagcga agtcggcgag atcctgcacg agggcatcaa cgaagccggg

26821 tcgttgggct cgttcatcgc ggccggcacc tcgtatcgca cgcacaacga accgatgatc
 26881 ccattttaca tcttctactc gatgttcggc ttccagcgca ccggcgatag ctcttggggc
 26941 gcggccgacc agatggctcg agggcttcgt ctccggggcca ccggccggcg caccaccttg
 5 27001 accggtgagg gcctgcaaca cggcgacggt cactcgttgc tgcctggccg caccaaoccg
 27061 gcggtgggtg cctacgaccc ggcttcggcc tacgaaatcg cctacatcgt ggaagcgga
 27121 ctggccagga tctgcgggga gaaccgggag aacatcttct tctacatcac cgtctacaac
 27181 gagccgtacg tgcagccgcc ggagccggag aacttcgato ccgagggcgt gctcgggggt
 27241 atctaccgct atcacggcg caccgagcaa cgcaccaaca aggcgcagat cctggcctcc
 10 27301 ggggtagcga tgcgccggcg cgtcggggca gcacagatgc tggccggcga gtgggatgct
 27361 gccgccgacg tctggctggt gaccagttgg ggcgagctaa accgcagcgg ggtggccatc
 27421 gagaccgaga agctccgcc cccgatcgg ccggcgggcg tgccctacgt gacgagagcg
 27481 ctggagaatg ctccggggccc ggtgatcggc gttcggact ggtgcgcgc ggtcccgag
 27541 cagatccgac cgtgggtgcc gggcacatac ctacgttgg gcaccgacgg gttcggtt
 15 27601 tcgcacactc ggcccggcg tcgcgctac ttcaacacg acgcgaatc ccaggtgttc
 27661 gcggttttgg aggcgttggc gggcgacggc gagatcgacc catcggttgc gttcggcgcc
 27721 gcccgccagt accggtatga cgacgtggcg cgtgcgccg agcagaccac aaatcccggt
 27781 cccggggcct aacgcggcg agccgaccgc cttggcgga atcttcaga aatctggcgt
 27841 agcttttagg agtgaacgac aatcagttgg ctccagtgc ccggccgagg tgcgcctcg
 20 27901 aactgctgga cactgtgccc gattcgctgc tgcggcggtt gaagcagtac tcggccggc
 27961 tggccaccga ggcagtttgc gccatgcaag aacggttgc gttcttcgcc gacatagaag
 28021 cgtccacgcg cggcagcgtg gcgctgggtg tgcagacggc cgtggtcaac ttctcgaat
 28081 ggatgcacga ccgcacagt gacgtcggt ataccgcga ggcattcgag ctggtgcccc
 28141 aggatctgac gcgacgcat gcgctgcgcc agaccgtgga catggtgcgg gtaaccatgg
 25 28201 agttctcga agaagtctgt ccctctctcg ccggttcga agagcagttg accgcctca
 28261 cgttggcat tttgaaatc agccgcgacc tggcattcac ccgccacag gcctacgccg
 28321 atcgcgcca ggacgaggc acctgggaca gccggatgga gccgacgctg tgggacggg
 28381 tggtaacgcg cgacaccggt ccgagctgc tgtccgggc ggccgcgctg aattgggaca
 28441 ccaccggccc ggcgaccgta ctggtgggaa ctccggcgcc cgttcacaa ggtccaaca
 30 28501 gcgacggcga cagcgagcgg gccagccagg atgtccgga caccgcggt cgccagggc
 28561 gcgtcgct gaccgacgtg cagggcacct ggctgttggc gatcgtctcc ggccagctgt
 28621 cgccaaccga gaagtctc aaagacctgc tggcagcatt cgccgacgc ccggttggtca
 28681 tcggcccaac gcgcccattg ctgaccggcg cgaccgcag cgctagcgag gcgatccg
 28741 ggatgaacgc cgtgccggcg tggcgggag ccggcgggc cgtgctggt agggaaatt
 35 28801 tggccgaacg cgccctgatg ggcgacgct ccgggatcgt ggccctgcat accgacgtga
 28861 tggggccct agccgatgcc ggaccgacg tcatcgagc gtagacga tatctggatt

28921 gtggcggcgc gattgaagct tgtgccagaa agttgttcgt tcatcaaac acagtgcggt
 28981 accggtctaa gcggatcacc gacttcaccg ggccgcgatcc caccagcca cgcgatggct
 29041 atgtccttcg ggtggcggcc accgtgggtc aactcaacta tccgacggc cactgaagca
 5 29101 tcgacagcaa tgcctgtgta tagattccct cgccggtcag aggggggtcca cgagggggcc
 29161 cggaaagata ccagggggcg cgtcggacgg aaagtgatcc agacaacagg tcgcggggacg
 29221 atctcaaaaa catagcttac agggccgttt tgttggttat atacaaaaac ctaagcagag
 29281 gticataatc tgttacaccg cgcaaaaccg tcttcacagt gtictcttag acacgtgatt
 29341 gcgttgcctc caccgggaca ggggttcgcaa accgagggaa tgttgtgcc gtgcttcag
 10 29401 ctgcccggcg cagcggacca gatcgcggcg tggtcgaaa ccgctgatct agatctgcc
 29461 cggctgggca ccaccggctc gaccgaggag atcaccgaca ccgcggtcgc ccagccattg
 29521 atcttcgccg cgactctgct gggccaccag gaactggcgc gccgatgctg gtcgccggc
 29581 aaggacgtca tcgtggccgg ccactccgtc ggcgaaatcg cgccctacgc aatcgccgt
 29641 gtgatagccg ccgacgacgc cgtcgcgctg gccgccacc ccgcccgcga gatggccaag
 15 29701 gcctgcgcca ccgagccgac cggcatgtct gcggtgctcg gcggcgacga gaccgaggtg
 29761 ctgactgcgc tcgagcagct cgacttggtc ccggcaaac gcaaccgcc cgccagatc
 29821 gtcgtgcgcc gccggtgac cgcgttggag aagctcgccg aagaccgcc ggccaaggcg
 29881 cgggtgctg cactgggtgt cgccggagcg ttccacaccg agttcatggc gcccgcaact
 29941 cgcgcttfg cggcgccgc gcccaaacat gcaaccgcgc accccaccgc cagctgctg
 20 30001 tccaaccgcg acgggaagcc ggtgacatcc gcggccgcgg cngtggacac cctgtctcc
 30061 cagctcacc ccaccggtgc atgggacctg tgcaccgcga cgtctgcgca acacacagt
 30121 accgctgatc tggagtccccc ccccgccggc acgottagcg gtatcgccaa acgcaact
 30181 cgggggggtc cggcacgcgc cgtcaagtca ccgcagacc tggacgagct ggcaaaccta
 30241 taaccgcgga ctgcggcaga acaaccacat accgtcagt tcatgtgta cacaacatat
 25 30301 tacgaaggga agcatgctgt gcctgtcact caggaaagaa tcatgcccgg tatgccgag
 30361 atatcgaag aggtaacccg tatcgagccg tccgagatca ccccgagaa gtctgtgtc
 30421 gacgacctgg acatcgactc gctgtgatg gtgcgagatc ccgtgcagac cgaggacaag
 30481 tacggcgta agatccccga cgaggacctc gccggtctgc gtaccgtcgg tgacgtgtc
 30541 gcctacatcc agaagctcga ggaagaaaa ccggaggccg ctcaggcgtt gcgcgcgaag
 30 30601 attagctcgg agaaccocga tgcgttgc aaacttcagg cgaggcttga gggcgagtcc
 30661 aagttagtca gccttcacc gctaattggc gtttccccg cgttgtgtg accgcggtca
 30721 cagcgacgac gtcgatctc ccggacatcg agagcacgtg gaagggtctg ttggccggcg
 30781 agagcggcat ccagcactc gaagacgagt tctcaccaa gtgggatcta gcggtcaaga
 30841 tcggcggtca ctcaaggat ccggtcgaca gccacatgg ccgactcgac atgcgacgca
 30901 tgtctactgt ccagcggatg ggcaagtgc tggcgccga gctatgggag tccgccggca
 35 30961 ccccgagggt cgatccagac cgggtccgc ttgtgtcgg caccggtcta ggtggagccg

31021 agaggattgt cgagagctac gacctgatga atgcgggcgg ccccccgaag gtgtccccc
 31081 tggccgttca gatgatcatg cccaacgttg ccgcggcggg gatccgtctg cagcttgggg
 31141 ccgcgcggg ggtgatgacc ccggtgtcgg cctgtctgc gggtcggaa gcgatcgccc
 5 31201 acgcgtggcg tcagatcgtg atggcgacg ccgacgtcgc cgtctcggc gggtcgaag
 31261 gacctatga ggcgctgcc atcgcggcgt tctcatgat gggggccatg tcgacccgca
 31321 acgacgagcc tgacggggcc tccggccgt tcgacaagga ccgcgacggc ttgtgtctg
 31381 gcgaggccgg tgcgctgatg ctatcgaga cggaggagca gcgcaagcc cgtggcgcca
 31441 agccgttggc ccgattctg gggtccgcta tcacctgga cgccttcat atgtggcg
 10 31501 ccgcggccga tgggttctg ccggtaggg cgatgactcg ctgcgtggag ctggccgggt
 31561 gtctcgccg ggacatcgac cagctcaac gcacggcac ggcgacgct atcgcgacg
 31621 ccgcggaggc caacgccatc cgcgtccg cgtgtgatca gcccgcgtg tacgcggca
 31681 agtctgcgt gggccactcg atcgccggc tgggtcgtc cgaagtcgtg ctacgggtg
 31741 tgacgtcgc cgacggcgic atccggcca cctgaacta cgagacacc gatccgga
 15 31801 tcgacctga cgtctgcgc gggaaccgc gctatggca ttaccgctac cgatcaaca
 31861 actcgttcg gttcggcgc cacaatggt cgttgcctt cggcggttac tgaagcaga
 31921 catcggggt cgcgagccc gaggtggggg tcccccgct tgcggggcg agtcgggacg
 31981 atatggaagg aacgttcga agacaatga cggagctggt taccgggaaa gccttccct
 32041 acgtatcgt caccggcatc gccatgacga ccgcctcgc gaccgacgc gaggactagt
 20 32101 ggaagtgtt gctgaccgc caaagcgga tccgtacgt cgaatgacca ttctcgagg
 32161 agtctgacct gccagtgc atcgccgac atcgtctga ggaattcgac caccagctga
 32221 ccgggatcga actgcgccg atgggatacc tgcagcggat gtccacctg ctgagccgc
 32281 gcctgtgga aatgccggc tcaccggagg tggacacaa tcgattgat gtgtccatg
 32341 gcaaccgctt ggttccgcc gaggaactgg tctcagtta cgacgatat cgcgtcgcg
 25 32401 gaatgaagg ggtctccgc ctgaccgtgc agaagtacat gcccaacggg gccgcgcgg
 32461 ccgtcgggtt ggaacggcac gccaaaggcc gggtgatgac gccgctatcg gcgtgcgat
 32521 ccgcgcggca ggcacgcgc cgtcgtcgc agcagatgt gctggagag gccgatccg
 32581 ccattcgcg cggcgtggag accaggatcg aagcgtgcc catcgccggg ttgcctcaga
 32641 tgcgatcgt gatgccacc aacaacgacg accccgccgt tgcgtccgc ccattcgaca
 30 32701 gggaccgca cggcttltg ttggcgagg gcggcgccct tctgtgatc gagaccagg
 32761 agcacgcaa ggcacgtggc gccaatcgc tggccggat catggcgcc agcatcact
 32821 ccgatggctt ccacatggt gccccggacc ccaacgggga acgcgcggc catgcaatta
 32881 ccggggcat tcagctggc ggccctgcc ccggcgacat cgaccagtc aatgcgcacg
 32941 ccaccgcac ccagctggc gacctggcc aagcgaggc catcaacaac gccttggcg
 33001 gcaaccgacc ggcgtgtgac gcccgaagt ctgcctcgg ccactcgtg ggcgcggtg
 35 33061 gcgcggctga atgatctg acgtgtcgc cgttcgcga tcaggtgat ccggccgac

33121 tgaatctggt aaacctcgat cccgagatcg atttggacgt ggtggcgggt gaaccgcgac
 33181 cgggcaatla ccggtatgcg atcaataact cgttcggatt cggcgggcac aacgtgggcaa
 33241 tgccttcgg acggtactaa accccagcgt tacgcgacag gagacctgcg atgacaatca
 5 33301 tggcccccga ggcgggtggc gagtgcctcg accccgcgca tccgtgtttg cggctgagca
 33361 acttcttga cgaaggcagc gtgggaattc tgcacgagcg tgaccgctcc ggagtgctgg
 33421 ccggcgccggg cacogtcaac ggtgtgcgca ccatcgcgtt ctgcaccgac ggcaccgtga
 33481 tggcgccggc catggcgctc gaggggtgca cgcacatcgt caacgcctac gacactgcca
 33541 tcgaagacca gagtccatc gtgggcattc ggcaticggg tgggtcccg cttgctgaag
 10 33601 gttgtcgggc gctgcacgcg gtaggccagg ttgtcgaagc catgatccgc gcgtccggct
 33661 acatcccgca gatctcgggt gtcgtcgggt tcggcccgcg cggcgccgac tacgagccgg
 33721 cgttgaccga cgtcgtcgtc atggcgccgg aaagccgggt gttcgtacc gggcccgacg
 33781 tgggtcgagc cgtcaccgcg gaggacgtcg acatggcctc gtcggtggg ccggagacc
 33841 accacaagaa gtccgggggtg tgccacatcg tcggcgacga cgaactcgat gctacgacc
 15 33901 gtggcgcccg gttgggtcgga ttgtctgcc agcaggggca tticgatcg agcaaggcgg
 33961 aggcgggtga caccgacatc cagcgctgc tggcggaatc ctgcgacgt cctacgacg
 34021 tgcgtccgat cgtgacggcg atcctcgatg cggacacacc gtgcgacgag ttacaggcca
 34081 attggcgccc gtcgatgggt gtcggcgctg gtcggctgct ggggtcgacg gttgggtgtac
 34141 tggccaacaa cccgtacgc ctggcgcggt cctgaactc cgaagcgca gagaaggcag
 20 34201 cgcgtttcgt cggcgtgtgc gacgcgttc ggaatccgt ggtgtgtgtg gtcatgtgc
 34261 cgggctatct gccgggtgtc gaaccagggt ggggtggcgt ggtgcgccgt ggccocaagt
 34321 tctgcacgc gttcggcgag tgcaccgttc cgggggtcac gctgttcacc cgaagacct
 34381 acggcggggc atacattgcg atgaactccc ggtcgttga cgcgaccaag gttgtccct
 34441 ggcggagcgc cgaagtcgcg gtgagggcg ctaaggcgcc cgtcggcatc ctgcacaaga
 25 34501 agaagttggc cgccgctccg gacgacgaac gcgaagcgct gcagcaccag ttggcgccgg
 34561 agcatgagcg catcggcgcg ggggtcgaca gtcgctgga catcggtgtg gtcgacgaga
 34621 agatcgacc gccgcatact cgcagcaagc lcaccgagcg gctgcgcgag gctcggcac
 34681 ggcgcggccg ccacaagaac atccgctgt agttctgacc gcgagcagac gcagaatcgc
 34741 acgcgcgagg tccgcgcgt gcgattctgc gttctctgc cagtatccc cagcggtgtg
 30 34801 tggtaacgc gaggcgtccc tcgatgctc ggacgggtgc taaccagcg ctaacaalc
 34861 tcgagaagcg cggcggttc gccaccacgc cgaattgct caggtctatg accgccaac
 34921 agctcgact ccaagtga aaacggcgcc tctctcgt ttgttacgg gttacgcgg
 34981 cacaagagcc ggacctgtt gccgcgttg cggctctga tgtgtcatg ggggggcacg
 35041 ccgtcgcgtg tctggcacc gccgccgct tgtatggatt cgacacgaa aacacgctc
 35101 ctatccatg gctcgatccc ggaglaagga tgcggccac gttcggtctg atgtccacc
 35161 aacgcgtcgg tggccgctc caacgggtgt caggtctctc cgcgaccgcg ccgcatgga

35221 ctgccgtgga ggtgcacga cagtgcgcc gccgcgggc gctgccacc ctgcagccg
 35281 cactacggctc aatgcgtgc gctgcagtg aaattgaaa cgccgttgct gagcagcgag
 35341 gccgccgagg catcgtcgc gcgcgcgaac tctacctt cgcgcagga cgcgcggaat
 5 35401 cggccatgga gagcgaggt cggctcgtca tgatcgacca cgggctcgcg tgcgccgaac
 35461 ttcaataccc gatacagggc cagcgtggcg aaatgtgcg agtcgactc gcctggccc
 35521 acatgcgtct cgcgccgaa tacganaagca tcgagtggca cgcgggaccg gcggagatgc
 35581 tgcgcgacaa gacacgtcgg gccaaagctcc aagagctcgg gtggagcatt gtcccgaattg
 35641 tegtgcagca tgtcagacgc gaaccggcc gcctgcggc ccgcatcgcc cgcacactcg
 10 35701 accgcgcgcg tatggccggc tgaccgtcgg tgagcagacg cagagatcgca ctgcggccgg
 35761 cgcagtgcga ctctcgtct gctcgcctc aacggctgag gaactccta gccacggcga
 35821 ctacgcctc cgcattccgt ggcaccagac cgaatccggg ccggcgctcg aggatctgt
 35881 ccacatccag cgcgccctca tgggtcaccg cgtattgaa ctccgcccg gtacgtcga
 35941 tgccgtcggc gaccggctc gtcggccgct cacatggcg gcggcgacgc acgttggccg
 15 36001 cctcgcgcc gtaccggcc accagcgact cgggcaatc ggccgccgat ccggggcgccg
 36061 gccagggtt cgcgggtgc ccgacagcg cagcgtgcg agtcggacg ttccggctc
 36121 gcaggtgcg cagcgtgatg gcgcgatca gcacatctc tgcattgtag cgtattccg
 36181 tcagctggc gccgaccaca ctgacacgc ccgacggcga ttcaaaaca cgcgtgtcac
 36241 gcgaagctc ggcgggtcgg cctggacac cagcaccgcc ggtgtcgatt agcggccgca
 20 36301 atccgcata ggcaccgatg acatcctgg tgccgaccgc cgtcccaat gcggtgttca
 36361 ccgatccag caggaaacgt atctctccg aagacggtg tggcacatcg ggaatcggc
 36421 cgggtgcgtc ttgtcgtgc agcccgagat agatccgcc cagctgctc ggcattggca
 36481 acacgaagc gttcagctca ccggggatcg gaatgtcag cgcggcagtc ggaattggca
 36541 acgactcgc gtcgaagacc agatgtgtgc cgcggctcgg gcgtagcctc agggacgggt
 25 36601 cgatctacc cgcacacgc ccgcgcggt tgatgacggc acgcgcgac agcgcgaacg
 36661 actgccgggt cgcgggtgc gtaactcca ccgaagtgc ggtgacatc gacgcgcca
 36721 cgtaatgag gatgcgggc ccgtgctgg ccgcggtgc cgcgacggcc atgaccagcc
 36781 gggcgtcgc gataattgc ccgtgtagc cagcagacc accgtcgagg ccgtccgcc
 36841 gaacgtggg agcaattcc accaccgtg acgccgggat tcggcgcat cggggcaacg
 30 36901 tcgcgcggc cgtaccgct agcaccgca aagctgcgc gccagagaaa ccggcagca
 36961 ccaacggcg cttggtgtga ccatcgacg gcaacaacg gaccagttgc ctgcgcggg
 37021 gcacgagatg agggagcttg cgtgtcatca ggattccgc ttgcagcgcg ctgcgcggg
 37081 cgatccacc gttgcgctg gccagatagc gcagaccgc gtgcaccaac ttgcagctc
 37141 agcgtgtgt gccgaacgc agatcatgct ttccacca ggcaccgtc agaccgcggg
 37201 tggcagcatc taaggcaatg ccaacaccg laattcgcc gcctatcac atgacgtcga
 37261 gtgcgccacc gtcggccagt gcgtcaggt cggcgagcg acgcgcgcg ttgagtgcg

37321 ccgagtgggg catcagcaca aataaccgtt cagtcgctgg gtaagtcgg tgccacgcgc
 37381 ggcggaatcg aggatcgaat cgcagatgtc gcgcgactgg atgtcgcagt ggcgcatcag
 37441 caacaccatg gtccccagtc gacgacgctc gccggagcgc acactgcccg accgctgcgc
 5 37501 cactgtcagc cggcgggcca acccctcgat caggacatgc tggctggctc cgaggcgtc
 37561 ggtgatgtac accctggcca gctccgagtg catgaccgac atgatcagat cgtcaccccg
 37621 caaccggctg gccaccgcga caatctgctt taccaacgct tcccggtcgt ccccgctgag
 37681 ggcgacctcc cgcagcacgt cggcgatatg gctgtgcagc atggacgcca tgatcgaccg
 37741 ggtgtccggc cagcgacggt alacggtcgg gcggctcacg cccgcgcgcg gggtgatctc
 10 37801 ggcaagtgct acccggcca cgcgtaatc gacgacgcag ctgcgcgtc cccgcaggat
 37861 acgaccaccg gtatccgcgc ggtcattact catggacagc atgtgtaala ctgtaacgcg
 37921 tgactaccg cgaggaaact ctccaccga tgaatggga cgcgtgggga gatcccgccg
 37981 cgcccaagcc actttctgat gggtccgggt cgtgtctgaa cgaggttggt ggctagccg
 38041 actcggagca gcccgaaact gacccgcgc aggtgcagct gcgccgcgc gccctgtcgc
 15 38101 gggcagacca (SEQ ID NO: 24)

5.9. X-linked Inhibitor of Apoptosis Protein ("XIAP")

GenBank Accession # U45880:

1 gaaaaggtgg acaagtccta tttcaagag aagatgactt ttaacagttt tgaaggatct
 20 61 aaaactgtg tacctgcaga catcaataag gaagaagaat ttgtagaaga gttaataaga
 121 ttaaaaactt ttgctaattt tccaagtggt agtctgttt cagcatcaac actgcacaga
 181 gcagggttct ttatactgg tgaaggagat accgtgcggt gctttagtgt tcatgcagct
 241 gtatagatag ggcaatatgg agactcagca gttggaagac acaggaaaat atocccaat
 301 tgcagattta tcaacggctt ttatctgaa aatagtcca cgcagtctac aaattctggt
 25 361 atccagaatg gtcagtacaa agttgaaac tatctgggaa gcagagatca ttftgcctta
 421 gacaggccat ctgagacaca tgcagactat ctfttgagaa ctgggcaggt tgtatataa
 481 tcagacacca tatacccgag gaacctcgcc atgtattgtg aagaagctag attaaagtc
 541 tticagaact ggccagacta tgcacaccta accccaagag agttagcaag tctcgactc
 601 tactacacag gtattgttga ccaagtgcag tgctttgtt gtgttggaat actgaaaaat
 30 661 tgggaacctt gtatcgtgc ctggtcagaa cacaggcgac acttctctaa ttgctcttt
 721 gtttgggcc ggaactctaa tattgaagt gaatctgatg ctgtgagtic tgataggat
 781 ttcccaatt caacaantct tcaagaant ccatccatgg cagattatga agcacggatc
 841 ttacttttg ggcacatgat atactcagti aacaaggagc agcttgcaag agcttgatt
 901 tatgctttag gtgaaggatg taaagtaag tgccttact gtggaggagg gctaactgat
 35 961 tggagccca gtgaagacc ttgggaacaa catgctaaat ggtatccagg gtcaaatat
 1021 ctgttagaac agaagggaca agaatatata aacaatatc atttaactca ttactgtgag

1081 gaggtctctgg taagaactac tgagaaaaca ccatcactaa ctagaagaat tgatgatacc
 1141 atcttccaaa atcctatgggt acaagaagct atacgaatgg gggtcagttt caaggacatt
 1201 aagaaaaataa tggaggaaaa aattcagata tctgggagca actataaatc acitgaggtt
 5 1261 ctggttgacg atctagttaa tgcctcagaaa gacagtatgc aagatgagtc aagtcagact
 1321 tcatatcaga aagagattag tactgaagag cagctaaagg gccctgaaga ggagaagctt
 1381 tgcaaaatct gtatggatag aaatattgct atcgttttg ttcttctgg acatctagtc
 1441 actgtataac aatgtctcga agcagttgac aagtgctcca tctgctacac agtcattact
 1501 ttcaagcaaa aaatttttat gtttlaatct aactctatag taggcatggt atgttgttct
 10 1561 tattacctg attgaatgtg tgatgtgaac tgactttaag taatcaggat tgaattccat
 1621 tagcatttgc taccaagtag gaaaaaaaat gtacatggca gtgttttagt tggcaatata
 1681 atctttgaat ttcttgaltt ttcagggtat tagctgtait atccatttt ttacttgta
 1741 ttaattgaa accatagact aagaataaga agcatcatat tataactgaa cacaatgtgt
 1801 attcatagta tactgaltta attictaagt gtaagtgaat taatcatctg gattttttat
 1861 tcttttcaga taggcttaac aaatgagct tctgtatat aaatgtggag attagagtta
 15 1921 atctcccaa tcacataatt tgttttgtg gaaaaaggaa taaattgttc catgctgggtg
 1981 gaaagataga gattgttttt agaggttgggt tgtgtgttt taggattctg tccattttct
 2041 tgtaaaggga taaacacgga cgtgtcggaa atatgttgt aaagtgttt gccattgttg
 2101 aaagcgtatt taatgataga atactatcga gccaacatgt actgacatgg aaagatgtca
 2161 gagatatgtt aagtgtaaaa tgcaagtggc gggacactat gfatagtctg agccagatca
 20 2221 aagtatgtat gttgttaata tgcatagaac gagagatttg gaaagatata caccaaactg
 2281 ttaattgttg ttctcttcg gggagggggg gattggggga ggggccccag aggggtttta
 2341 gaggggctt ttactttcg actttttca ttgttctg ttggatttt ttalaagtat
 2401 gtagacccc aagggtttta tgggaactaa caicagtaac ctaaccccc tgactatcct
 2461 gtgctcttc tagggagctg tgtgttttc caccaccac ccttccctct gaacaaatgc
 25 2521 ctgagtctg gggcactttg (SEQ ID NO: 25)

General Target Region:

Internal Ribosome Entry Site (IRES) in 5' untranslated region:
 30 5'AGCUCCUAUAACAAAAGUCUGUUGCUUGUGUUUCACAUUUUGGAUUU
 CCUAAUAUAUAUGUUCUUUUUAGAAAAGGUGGACAAGUCCUAUUUUU
 AAGAGAAG3' (SEQ ID NO: 26)

Initial Specific Target Motif:

RNP core binding site within XIAP IRES
 35 5'GGAUUUCCUAUAUAUAUAUGUUCUUUUU3' (SEQ ID NO: 27)

5.10. Survivin

GenBank Accession # NM_001168:

1 ccgcagatt tgaatcgcg gacccgttgg cagagggtggc ggcggcgcca tgggtgcccc
 5 61 gacgttcccc cctgcctggc agccctttct caaggaccac cgcattctcta catcaagaa
 121 ctggcccttc ttggagggtc gcgcctgcac cccggagcgg atggccgagg ctgcttcatt
 181 ccactgcccc actgagaacg agccagactt ggcacagtgt ttctctgct tcaaggagct
 241 ggaaggctgg gagccagatg acgaccccat agaggaacat aaaaagcatt cgtccggttg
 301 cgcttctcti tctgtcaaga agcagtttga agaattaacc ctgggtgaat ttitgaaact
 10 361 ggacagagaa agagccaaga acaaaattgc aaaggaaacc aacaataaga agaaagaatt
 421 tgaggaaact gcgaagaag tgcgccgtgc catcgagcag ctggctgcca tggattgagg
 481 cctctggccg gagctgcctg gtcccagagt ggctgcacca cticcagggt ttatccctg
 541 gtgccaccag cctctctgtg ggcctcttag caatgtctta ggaaggaga tcaacattt
 601 caaatagat gtttcaactg tgcctctgtt ttgtctttaa agtggcacca gagggtgcttc
 15 661 tgctgtgca gcggggtgct ctggtaacag tggctgcttc tctctctc tctcttttt
 721 gggggctcat tttgtgtt ttgattccc ggcttaccag gtgagaagtg aggggagaa
 781 aaggcaggtg ccttttgcct agagctgaca gctttgttc cgtgggcaga gccttccaca
 841 gtgaatgtgt ctggacctca tttgtgtgag gctgtcacag tctgtagtgt ggacttgcca
 901 ggtgcctgtt gaattctgag tgcaggttcc ttatctgca cacctgtgcc tctctagagg
 20 961 acagtttttt tgtgtgtgt ttttttgt tttttttt ggtatagca tgaattgtgt
 1021 gtgatgagag aatggagaca gagtccctgg ctctctact gtttaacaac atgctttct
 1081 tatttgttt gaattgttaa ttacagaa agcacaaact acaattaaaa ctaagcacaa
 1141 agccattccla agtcattggg gaaacggggt gaacttcagg tggatgagga gacagaatag
 1201 agtgatagga agcgtctggc agatactcct ttgccactg ctgtgtgatt agacaggccc
 25 1261 agtgagccgc ggggcacatg ctggccgctc ctccctcaga aaaaggcagt ggcctaaatc
 1321 cttttaaat gacttgctc gatgtgtgg gggactgctt ggcctgctgc aggccgtgtg
 1381 tctgtcagcc caacttcac atctgtcac ttctccacac gggggagaga cgcagtcgcc
 1441 ccaggtcccc gcttttttg gaggcagcag ctcccgagg gctgaagtct ggcgtaaat
 1501 gatggatttg attcccccct ctccctgtca tagagctgca ggggtgattg ttacagcttc
 30 1561 gctggaacc tctggaggtc atctggctg ttcttgagaa ataaaagcc tgcatttc (SEQ ID NO: 28)

The present invention is not to be limited in scope by the specific
 embodiments described herein. Indeed, various modifications of the invention in addition
 to those described will become apparent to those skilled in the art from the foregoing
 35 description and accompanying figures. Such modifications are intended to fall within the
 scope of the appended claims.

Various publications are cited herein, the disclosures of which are incorporated by reference in their entireties.

5

10

15

20

25

30

35

The invention can be illustrated by the following embodiments enumerated in the numbered paragraphs that follow:

- 5 1. A method for identifying a test compound that binds to a target RNA molecule, comprising the steps of (a) contacting a detectably labeled target RNA molecule with a library of solid support-attached test compounds under conditions that permit direct binding of the labeled target RNA to a member of the library of solid support-attached test compounds so that a detectably labeled target RNA:support-attached test compound complex is formed; (b) separating the detectably labeled target RNA:support-attached test compound
10 complex formed in step (a) from uncomplexed target RNA molecules and test compounds, and (c) determining a structure of the test compound of the RNA:support-attached test compound complex.
- 15 2. The method of paragraph 1 in which the target RNA molecule contains an HIV TAR element, internal ribosome entry site, "slippery site", instability element, or adenylate uridylate-rich element.
- 20 3. The method of paragraph 1 in which the RNA molecule is an element derived from the mRNA for is tumor necrosis factor alpha ("TNF- α "), granulocyte-macrophage colony stimulating factor ("GM-CSF"), interleukin 2 ("IL-2"), interleukin 6 ("IL-6"), vascular endothelial growth factor ("VEGF"), human immunodeficiency virus I ("HIV-1"), hepatitis C virus ("HCV" - genotypes 1a & 1b), ribonuclease P RNA ("RNaseP"), X-linked inhibitor of apoptosis protein ("XIAP"), or survivin.
- 25 4. The method of paragraph 1 in which the detectably labeled RNA is labeled with a fluorescent dye, phosphorescent dye, ultraviolet dye, infrared dye, visible dye, radiolabel, enzyme, spectroscopic colorimetric label, affinity tag, or nanoparticle.
- 30 5. The method of paragraph 1 in which the test compound is selected from a combinatorial library comprising peptoids; random bio-oligomers; diversomers such as hydantoins, benzodiazepines and dipeptides; vinyllogous polypeptides; nonpeptidal peptidomimetics; oligocarbamates; peptidyl phosphonates; peptide nucleic acid libraries; antibody libraries; carbohydrate libraries; and small organic molecule libraries including, but
35 not limited to, benzodiazepines, isoprenoids, thiazolidinones, metathiazanones, pyrrolidines, morpholino compounds, or diazepindiones.

6. The method of paragraph 1 in which screening a library of test compounds preferably comprises contacting the test compound with the target nucleic acid in the presence of an aqueous solution, the aqueous solution comprising a buffer and a combination of salts, preferably approximating or mimicking physiologic conditions

7. The method of paragraph 6 in which the aqueous solution optionally further comprises non-specific nucleic acids comprising DNA, yeast tRNA, salmon sperm DNA, homoribopolymers, and nonspecific RNA.

8. The method of paragraph 6 in which the aqueous solution further comprises a buffer, a combination of salts, and optionally, a detergent or a surfactant. In another embodiment, the aqueous solution further comprises a combination of salts, from about 0 mM to about 100 mM KCl, from about 0 mM to about 1 M NaCl, and from about 0 mM to about 200 mM MgCl₂. In a preferred embodiment, the combination of salts is about 100 mM KCl, 500 mM NaCl, and 10 mM MgCl₂. In another embodiment, the solution optionally comprises from about 0.01% to about 0.5% (w/v) of a detergent or a surfactant.

9. Any method that detects an altered physical property of a target nucleic acid complexed to a test compound attached to a solid support from the unbound target nucleic acid may be used for separation of the complexed and non-complexed target nucleic acids in the method of paragraph 1. Methods such as flow cytometry, affinity chromatography, manual batch mode separation, suspension of beads in electric fields, and microwave are used for the separation of the complexed and non-complexed target nucleic acids.

10. The structure of the substantially one type of test compound of the RNA:test compound complex of paragraph 1 is determined, in part, by the type of library of test compounds. In a preferred embodiment wherein the combinatorial libraries are small organic molecule libraries, mass spectroscopy, NMR, or vibration spectroscopy are used to determine the structure of the test compounds. In an embodiment wherein the combinatorial libraries are peptide or peptide-based libraries, Edman degradation is used to determine the structure of the test compounds.

WHAT IS CLAIMED IS:

1. A method for identifying a test compound that binds to a target RNA molecule, comprising the steps of:
- (a) contacting a detectably labeled target RNA molecule with a library of solid support-attached test compounds under conditions that permit direct binding of the labeled target RNA to a member of the library of solid support-attached test compounds so that a detectably labeled target RNA:support-attached test compound complex is formed;
 - (b) separating the detectably labeled target RNA:support-attached test compound complex formed in step (a) from uncomplexed target RNA molecules and test compounds by flow cytometry; and
 - (c) determining a structure of the substantially one type of test compound of the RNA:support-attached test compound complex by mass spectroscopy.

SEQUENCE LISTING

<110> PCT Therapeutics, Inc.

<120> METHODS FOR IDENTIFYING SMALL MOLECULES THAT BIND SPECIFIC RNA
STRUCTURAL MOTIFS

<130> 10589-008-228

<140> To be assigned

<141> 2002-04-11

<150> 60/282,966

<151> 2001-04-11

<160> 28

<170> PatentIn version 3.0

<210> 1

<211> 21

<212> RNA

<213> Homo sapiens

<400> 1
auuuuuuuau uuauuuuuu a

21

<210> 2

<211> 17

<212> RNA

<213> Homo sapiens

<400> 2
auuuuuuuau uuauuuu

17

<210>	3	
<211>	15	
<212>	RNA	
<213>	Homo sapiens	
<400>	3	15
	wauuuuuuuu uuuaaw	
<210>	4	
<211>	13	
<212>	RNA	
<213>	Homo sapiens	
<400>	4	13
	wwauuuuuuu aww	
<210>	5	
<211>	13	
<212>	RNA	
<213>	Homo sapiens	
<400>	5	13
	wwwauuuuuw www	
<210>	6	
<211>	1643	
<212>	DNA	
<213>	Homo sapiens	
<400>	6	60
	gcagagagacc agctaagagg gagagaagca actacagacc cccoctgaaa acaacctca	
	gacgccacat cccctgacaa gctgccaggc aggttctctt cctctcacat actgacctac	120
	ggctccaccc tctctccct ggaaggaca ccatgagcac tgaagcatg atccgggagc	180
	tggagctggc cgaggaggcg ctccccaaga agacaggggg gccccagggc tccaggcggc	240
	gcttcttct cagcctcttc tcttctctga tcgtggcagg cgccaccacg ctcttctgcc	300

```

tgctgcactt tggagtgate ggccccaga ggaagagtt cccaggagac ctctctctaa 360
tcagccctct gggccaggca gtcagatcat ctctctogaa cccagagtgaac aagcctgtag 420
cccatgttgt agcaaacctt caagctgagg ggcagctcca gtggctgaac cggcgggcca 480
atgccctcct ggccaatggc gtggagctga gagataacca gctgggtggtg ccactagagg 540
gcctgtacct catctaetcc caggctctct tcaaggcca aggotgcccc tccaccctatg 600
tgctctctac ccacaccatc agccgcatcg ccgtctctca ccagaccaag gtcaacctcc 660
tctctgccat caagagcccc tgccagaggg agaaccoca gggggctgag gccaaagccct 720
ggtagtagcc catctatctg ggaggggtct tccagctgga gaaggggtgac cgactcagcg 780
ctgagatcaa tcggcccagc tatctcgact ttgocgagtc tgggcaggtc tactttggga 840
tcattgcctt gtgaggaggga cgaacatcca accttcccaa acgctctccc tgccccaatc 900
cctttattac cccctccttc agacaccctc aacctcttct ggctcaaaaa gagaattggg 960
ggcttagggg cggaaaccaa gcttagaact ttaagcaaca agaccaccac ttogaaacct 1020
gggattcagg aatgtgtggc ctgcaagtg aattgctggc aaccactaag aattcaaaact 1080
ggggcctoca gaactcactg gggcctacag ctttgatccc tgacatctgg aatctggaga 1140
ccaggagacc tttggtcttg gccagatgc tgcaggactt gagaagacct caoctagaaa 1200
ttgacacaag tggaccttag gccctctctc ctocagatgt ttocagactt ccttgagaca 1260
cggagcccag cccctcccat ggagccagct cctctatatt atgtttgac ttgtgattat 1320
ttattattta tttattattt attattttac agatgaatgt atttatttgg gagaocgggg 1380
tatcctgggg gacccaatgt aggagctgcc ttggctcaga catgttttcc gtgaaaacgg 1440
agctgaacaa taggctgttc ccatgtagcc cctggcctc tgtgccttct ttgtattatg 1500
ttttttaaaa tatttatctg attaagttgt ctaaacatg ctgatttggg gaccaactgt 1560
cactcattgc tgagcctctg ctccccaggg gagttgtgtc tgtaatcgcc ctactattca 1620
gtggcgagaa ataaagtttg ott 1643

```

<210> 7

<211> 756

<212> DNA

<213> Homo sapiens

<400> 7

```

gctggaggat gggctgcag agcctgctgc tcttgggac tgtggcctgc agcatctctg 60
caccgcccg ctgcgccagc ccagcagcg agcctggga gcatgtgaat gccatccagg 120
aggccggcg tctctggaac ctgagtagag aactgctgc tgagatgaat gaaacagtag 180

```

```

aagtcatctc agaaatgttt gacctccagg agccgacctg cctacagacc cgccctggagc 240
tgtacaagca gggcctgcgg ggcagcctca ccaagctcaa gggccccttg accatgatgg 300
ccagccaacta caagcagcac tgcctcccaa ccccggaaac ttccctgtgca acccagacta 360
tcacctttga aagtttcaaa gagaacctga aggactttct gcttgtcacc ccctttgact 420
gctgggagcc agtccaggag tgagaccggc cagatgaggc tggccaagcc ggggagctgc 480
tctctcatga aacaagagct agaaactcag gatggtcacc ttggagggac caaggggtgg 540
gccacagcca tgggtgggag ggcctggacc tgcctgggc cactagacc ctgatacagg 600
catggcgcaa gaatgggaat attttatact gacagaaatc agtaatatat atatatattat 660
atttttaaaa tattttatta tttatttatt taagttcata ttccatatat attcaagatg 720
ttttaccgta ataattatta ttaaaaatat gcttct 756

```

<210> 8

<211> 756

<212> DNA

<213> Homo sapiens

```

<400> 8
tctggaggat gtggtgcgag agcctgctgc tcttgggcac tgtggcctgc agcatctctg 60
caccgcccg ctcgccagc cccagcacgc agccctggga gcatgtgaat gccatccagg 120
aggcccgcg tctcctgaac ctgagtagag acaactgctgc tgagatgaat gaaacagtag 180
aagtcatctc agaaatgttt gacctccagg agccgacctg cctacagacc cgccctggagc 240
tgtacaagca gggcctgcgg ggcagcctca ccaagctcaa gggccccttg accatgatgg 300
ccagccaacta caagcagcac tgcctcccaa ccccggaaac ttccctgtgca acccagacta 360
tcacctttga aagtttcaaa gagaacctga aggactttct gcttgtcacc ccctttgact 420
gctgggagcc agtccaggag tgagaccggc cagatgaggc tggccaagcc ggggagctgc 480
tctctcatga aacaagagct agaaactcag gatggtcacc ttggagggac caaggggtgg 540
gccacagcca tgggtgggag ggcctggacc tgcctgggc cactagacc ctgatacagg 600
catggcgcaa gaatgggaat attttatact gacagaaatc agtaatatat atatatattat 660
atttttaaaa tattttatta tttatttatt taagttcata ttccatatat attcaagatg 720
ttttaccgta ataattatta ttaaaaatat gcttct 756

```

<210> 9

<211> 825

<212> DNA

<213> Homo sapiens

```

<400>  9
atcactctct ttaatcacta ctcacattaa cctcaactcc tgcacacaatg tacaggatgc      60
aactcctgtc ttgcattgca ctaattcttg cacttgctac aaacagtgoa cctacttcaa      120
gttcgacaaa gaaaacaaag aaaacacagc tacaactgga gcatttactg ctggatttac      180
agatgatttt gaatggaatt aataattaca agaatcccaa actcaccagg atgotcacat      240
ttaagtttta catgccaag aaggccacag aactgaaaca gottcagtgt ctagaagaag      300
aactcaaac cctggaggaa gtgctgaatt tagotcaaag caaaaacttt cacttaagac      360
ccagggaactt aatcagcaat atcaacgtaa tagttctgga actaaaggga tctgaacaa      420
cattcatgtg tgaatatgca gatgagacag caaccattgt agaatttctg aacagatgga      480
ttacottttg tcaaacgcat atctcaacac taacttgata attaagtgot tccocactaa      540
aacatatcag gctctctatt tattttatta aatatttaa ttttatattt attgttgaat      600
gtatggttgc tacctattgt aactattatt cttaacttta aaactataaa tatggatctt      660
ttatgattct ttttgtaagc cctagggggt ctaaaatggt ttacottatt tatcccaaaa      720
atatttatta ttatgttgaa tgttaaatat agtatctatg tagattgggt agtaaaacta      780
tttaataaat ttgataata taaaaaaaaa aaacaaaaaa aaaaa      825

```

<210> 10

<211> 15

<212> RNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> (1)..(1)

<223> N = A, U, G, OR C

<220>

<221> misc_feature

<222> (15)..(15)

<223> N = A, U, G, OR C

<400> 10
 nauuuuuuuu uuuuu 15

<210> 11
 <211> 1125
 <212> DNA
 <213> Homo sapiens

<400> 11
 ttctgcccctc gagccaccgc ggaacgaaag agaagctcta tctgcctcc aggagcccag 60
 ctatgaactc cttctccaca agcgccttcg gtccagttgc cttctccctg ggggtgctcc 120
 tgggtgttgc tgctgccttc cctgcccag tacccccag agaagattcc aaagatgtag 180
 ccgccccaca cagacagcca ctcacctctt cagaacgaat tgacaaacaa attcgggtaca 240
 tctctgacgg catctcagcc ctgagaaagg agacatgtaa caagagtaac atgtgtgaaa 300
 gcagcaaaga ggcactggca gaaaacaacc tgaaccttcc aaagatggct gaaaaagatg 360
 gatgcttoca atctggattc aatgaggaga cttgcctggg gaaaatcctc actggtcttt 420
 tggagtttga ggtataccta ggtacctcc agaacagatt tgagagtagt gaggaacaag 480
 ccagagctgt gcagatgagt acaaaagtcc tgatccagtt cctgcagaaa aaggcaaaga 540
 atctagatgc aataaccacc cctgacccaa ccacaaatgc cagcctgctg acgaagctgc 600
 aggcacagaa ccagtggtgc caggacatga caactcatct cattctgcgc agctttaagg 660
 agttcctgca gtccagcctg agggctcttc ggcaaatgta gcattggcac ctcagattgt 720
 tgttgttaat gggcattcct tcttctggtc agaaacctgt ccactgggca cagaacttat 780
 gttgttctct atggagaact aaaagtatga gcgttaggac actattttaa ttatttttaa 840
 tttattaata tttaaatatg tgaagctgag ttaatttatg taagtcatat ttatatTTTT 900
 aagaagtacc aottgaaaca ttttatgtat tagttttgaa ataataatgg aaagtggcta 960
 tgcagtttga atatccttgg tttcagagcc agatcatttc ttggaaagt taggcttacc 1020
 tcaataaataa ggctaactta tacatatTTT taaagaaata tttatatTgt atttatataa 1080
 tgtataaatg gttttttatac caataaatgg catttttaaa aattc 1125

<210> 12
 <211> 3166
 <212> DNA
 <213> Homo sapiens

<400> 12
 aagagctcca gagagaagtc gaggaagaga gagacggggt cagagagagc gcgcgggcgt 60
 gcgagcagcg aaagcgacag gggcaaagtg agtgacctgc ttttgggggt gaccgccgga 120
 gcgcggcggtg agccctcccc cttgggatcc cgcagctgac cagtgcgcgt gacggacaga 180
 cagacagaca cgcgccccag ccccaagttac cacctcctcc ccggccgcgc gcggacagtg 240
 gacgcggcgc cgagccgcgc gcagggggcg gagccgcgcc ccggaggcgc ggtggagggg 300
 gtcgagctc gcgcgctgc actgaaactt ttcgtccaac ttctgggctg ttctcgttc 360
 ggaggagccg tggtcgcgc gggggaagcc gagccgagcg gagccgcgag aagtgtctagc 420
 tcgggcggg aggagccgca gccggaggag ggggaggagg aagaagagaa ggaagaggag 480
 aggggcgcg agtggcgact cggcgctcgc aagccgggct catggacggg tgagcggcg 540
 gtgtgcgag acagtgtcc agcgcgcgc cccccagcc ctggcccggc ctgcggccgc 600
 gaggaagagt agctgcgca ggcgcgagc agagcgggcc gccccacagc ccgagccgga 660
 gagggagcgc agccgcgcgc ccgggtcggg cctccgaaac catgaacttt ctgctgttct 720
 ggggtgcatg gagccttgcc ttgctgtctt acctccacca tgccaagtgg tcccaggctg 780
 ccccatggc agaaggagga gggcagaatc atcacgaagt ggtgaagttc atggatgtct 840
 atcagcgag ctactgccat ccaatcgaga cctgtgtgga catcttcagc gagtacctg 900
 atgagatcga gtacatcttc aagccatcct gtgtgcccct gatgcgatgc gggggtgtct 960
 ccaatgacga gggcctggag tgtgtgccca ctgaggagtc caacatcacc atgcagatta 1020
 tcgggatcaa acctcaccaa ggccagcaca taggagagat gagcttccca cagcacaaca 1080
 aatgtgaatg cagaccaaa aaagatagag caagacaaga aaatccctgt gggccttgc 1140
 cagagcggag aaagcatttg tttgtacaag atccgcagac gtgtaaatgt tcctgcaaaa 1200
 acacacactc gcgttgcaag gcgaggcagc ttgagttaaa cgaacgtact tgcagatgtg 1260
 acaagccgag gcggtgagcc gggcaggagg aaggagcctc cctcagggtt tcgggaacca 1320
 gatctctctc caggaaagac tgatacagaa cgatcgatac agaaaaccag ctgcgcgcac 1380
 cacaccatca ccatcgacag aacagtcctt aatccagaaa cctgaaatga aggaagagga 1440
 gactctgcgc agagcacttt gggtcgggag ggcgagactc cggcgggaagc attcccgggc 1500
 gggtgaccca gcacggtccc tcttggaatt ggattcgcca ttttattttt cttgctgcta 1560
 aatcaccgag ccggaagat tagagagttt tttttctggg attcctgtag acacaccac 1620
 ccacatacat acatttatat atatatatat tatatatata taaaaataaa tatctctatt 1680
 ttatatatat aaaatatata tattcttttt ttaaattaac agtgctaagt ttattggtgt 1740
 cttaactgga tgtatttgac tgctgtggac ttgagttggg aggggaatgt tccactcag 1800

```

atcctgacag ggaagaggag gagatgagag actctggcat gatotTTTTt ttgtcccact 1860
tggtggggcc agggctcctc cccctgccca agaattgca aggccagggc atgggggcaa 1920
atatgacca gttttgggaa caccgacaaa cccagccctg gcgctgagcc tctctacccc 1980
aggtcagacg gacagaaaga caaatcacag gtccgggat gaggcacccg gctctgacca 2040
ggagtttggg gagcttcagg acattgctgt gctttgggga ttccctccac atgtgcacg 2100
cgcatctcgc cccaggggac actgcctgga agattcagga gcctgggcgg ccttgcetta 2160
ctctcaactg ctcttgagt ttccaggagg ccaactggag atgtcccgcc gaagagaaga 2220
gacacattgt tgaagaagc agcccatgac agcgccctt cctgggactc gccctcatcc 2280
tcttctgtct ccccttctcg ggggtgcagc taaaaggacc tatgtctca caccattgaa 2340
accaatagtt ctgtccccc aggaacactg gttgtgtgtg tgtgagtggg tgaccttcc 2400
ccatccctcg gtcccttcc  tcccttccg aggcacagag agacagggca ggatccacgt 2460
gcccatgtg gaggcagaga aaagagaaag tgttttatat acggtactta ttaatatcc 2520
ctttttaatt agaaataga acagttaatt taattaaaga gtagggtttt ttttcagtat 2580
tcttggttaa tatttaatt caactattta tgagatgtat cttttgtctc ctcttgcctc 2640
cttatttgta ccggtttttg tatataaaat tcatgtttcc aatctctctc tccctgatcg 2700
gtgacagtca ctagcttacc ttgaacagat atttaatttt gctaacaactc agctctgccc 2760
tcccgcagcc cctggtctcc cagcacacat tcccttgaaa gagggtttca atatacatct 2820
acatactata tatatattgg gcaacttgta tttgtgtgta tatatatata tatatgttta 2880
tgtatatatg tgatcctgaa aaaataaaca tcgctattct gttttttata tgttcaaacc 2940
aaacaagaaa aaatagagaa ttctacatac taaatctctc tcttttttta attttaatat 3000
ttgttatcat ttattttatt gtgctactgt ttatccgtaa taattgtggg gaaaagatat 3060
taacatcacg tctttgtctc tagtgcagtt ttccgagata ttccgtagta catatttatt 3120
tttaacaac gacaagagaa tacagatata tcttaaaaa aaaaaa 3166

```

<210> 13

<211> 249

<212> RNA

<213> Homo sapiens

<400> 13

```

ccgggcuccau ggacggguga ggcggcggug ucgcgagaca gugcuccagc gcgcgcgcuc 60
cccgcccuug gcccgccuc gggccgggag gaagaguagc ucgccagggc gcccgaggaga 120
gcgggcgcgc ccacagcccg agccggagag ggacgcgagc cgcgcgcccc ggucggggcu 180

```

ccgaaaccacg gaacuuucug cugucuuuggg ugcauuggag ccuugccuug cugcucuacc 240
uccaccaug 249

<210> 14

<211> 9181

<212> DNA

<213> Homo sapiens

<400> 14
ggtctctctg gttagaccag atctgagcct gggagctctc tggctaacta gggaaaccac 60
tgcttaagcc toaataaagc ttgccttgag tgcttcaagt agtgtgtgcc cgtctgttgt 120
gtgaactctg taactagaga tccctcagac ccttttagtc agtgtggaaa atctctagca 180
gtggcgcccg aacaggagcc tgaagcgaa agggaaacca gagagctctc ctgcagcgag 240
gaactcggtt gctgaagcgc gcaogcgaa aggcgagggg cggcgacttg tgaglacgcc 300
aaaaattttg actagcggag gctagaagga gagagatggg tgcgagagcg tcagtattaa 360
gcgggggaga attagatcga tgggaaaaaa ttcggttaag gccaggggga aagaaaaaat 420
ataaattaaa acatatagta tgggcaagca gggagctaga acgattcgca gttaatcctg 480
gcctgttaga aacatcagaa ggctgtagac aaatactggg acagctacaa coactccctc 540
agacaggatc agaagaactt agatcattat ataatacagt agcaaccctc tattgtgtgc 600
atcaaaggat agagataaaa gacaccaagg aagctttaga caagatagag gaagagcaaa 660
acaaaagtaa gaaaaaagca cagcaagcag cagctgacac aggacacagc aatcaggtca 720
gccaaaatta ccttatagtg cagaacatcc aggggcaaat ggtacatcag gccatatcac 780
ctagaacttt aatgcactgg gtaaaagtag tagaagagaa ggctttcagc ccagaagtga 840
taoccatggt ttacgacatc tcagaaggag ccaaccacac agattttaac accatgctaa 900
acacagtggt gggacatcaa gcagccatgc aaatgttaaa agagaccatc aatgaggaag 960
ctgcagaatg ggatagatg catccagtgc atgcagggcc tattgcacca gccacagatga 1020
gagaaccaag ggaagtgc atagcaggaa ctactagtac ccttcaggaa caaataggat 1080
ggatgacaaa taatccacct atcccagtag gagaaattta taaaagatgg ataactctgg 1140
gattaaataa aatagtaaga atgtatagcc ctaccagcat tctggacata agacaaggac 1200
caagggaacc ctttagagac tatgtagacc ggttctataa aactctaaga gccgagcaag 1260
cttcacagga ggtaaaaaat tggatgacag aaacctgtgt ggtccaaaat gcgaaccag 1320
attgtaagac tatttttaaa gcattgggac cagcggctac actagaagaa atgatgacag 1380
catgtcaggg agtaggagga cccggccata aggcgaagat tttggctgaa gcaatgagcc 1440


```

aagtaacaaa ttcagctacc ataatgatgc agagaggcaa ttttaggaac caaagaaga 1500
ttgttaagtg ttccaattgt ggcaagaag ggcacacagc cagaaattgc agggccocta 1560
ggaaaaaggg ctgttggaag tgtggaaagg aaggacacca aatgaaagat tgtaactgaga 1620
gacaggctaa ttttttaggg aagatctggc ctctctacaa ggggaaggcca ggggaatttc 1680
ttcagagcag accagagcca acagccccc cagaagagag cttcaggtct ggggtagaga 1740
caacaactcc cctcagaag caggagccga tagacaagga actgtatcct ttaacttccc 1800
tcagggtact ctltggcaac gacccctcgt cacaataaag ataggggggc aactaaagga 1860
agctctatta gatacaggag cagatgatac agtattagaa gaaatgagtt tgccagggaag 1920
atggaacca aaaaatgatag ggggaattgg aggttttata aaagtaagac agtatgatca 1980
gatactcata gaaatctgtg gacataaagc tatagggtaca gtattagtag gacctacacc 2040
tgtaacata attggaagaa atctgttgac tcagattggg tgcaatttaa attttcccat 2100
tagccctatt gagactgtac cagtaaaatt aaagccagga atggatggcc caaagttaa 2160
acaatggcca ttgacagaag aaaaaataa agcattagta gaaatttgta cagagatgga 2220
aaaggaaggg aaaatttcaa aaattgggcc tgaaaatcca tacaatactc cagtatttgc 2280
cataagaaa aaagacagta ctaaattggag aaaattagta gatttcagag aacttaataa 2340
gagaactcaa gacttctggg aagtccaatt aggaatacca catcccgagc ggttaaaaaa 2400
gaaaaatca gtaacagtac tggatgtggg tgatgcatat ttttcagttc ccttagatga 2460
agacttcagg aagtatactg catttaccat acctagtata acaatgaga caccagggat 2520
tagatatcag tacaatgtgc ttccacaggg atggaagaa tcaccagcaa tattccaaag 2580
tagcatgaca aaaaacttag agccttttag aaaacaaaat ccagacatag ttatctatca 2640
atacatggat gatttgtatg taggatctga cttagaaaata gggcagcata gaacaaaaat 2700
agaggagctg agacaacatc tgttgagggt gggacttacc accaccagaca aaaaacatca 2760
gaaagaacct ccattccttt ggatgggtta tgaactccat cctgataaat ggacagtaca 2820
gcctatagtg ctgccagaaa aagacagctg gactgtcaat gacatacaga agttagtggg 2880
gaaattgaat tgggcaagtc agatttaccc agggattaaa gtaaggcaat tatgtaaact 2940
ccttagagga accaaagcac taacagaagt aataccacta acagaagaag cagagctaga 3000
actggcgcaa aacagagaga ttctaaaaga accagtacat ggagtgtatt atgaccatc 3060
aaaagactta atagcagaaa tacagaagca ggggcaaggc caatggacat atcaaattta 3120
tcaagagcca tttaaaaatc tgaaaacagg aaaatattgca agaattaggg gtgccacac 3180
taatgatgta aaacaattaa cagaggcagt gcaaaaaata accacagaaa gcatagtaat 3240
atggggaag actcctaact ttaactgcc catacaaaag gaaacatggg aaacatggtg 3300
gacagagtat tggcaagcca cctggattcc tgagtgggag tttgttaata cccctccctt 3360

```

```

agtgaatta tggtagcagt tagagaaaga acccatagta ggagcagaaa ccttctatgt 3420
agatggggca gctaacaggg agactaaatt aggaaaagca ggatatgtta ctaatagagg 3480
aagacaaaaa gttgtcaccg taactgacac aacaaatcag aagactgagt tacaagcaat 3540
ttatctagct ttgcaggatt cgggattaga agtaaacata gtaacagact cacaatatgc 3600
attaggaatc attcaagcac aaccagatca aagtgaatca gagttagtca atcaataat 3660
agagcagtta ataaaaaagg aaaaggtcta tctggcatgg gtaccagcac acaaagggaat 3720
tggaggaaat gaacaagtag ataaattagt cagtgtctga atcaggaaag tactattttt 3780
agatggaata gataaggccc aagatgaaca tgagaaatat cacagtaatt ggagagcaat 3840
ggctagtgat tttaacctgc cacctgtagt agcaaaagaa atagtagcca gctgtgataa 3900
atgtcagcta aaaggagaag ccatgcatgg acaagtagac tgtagtccag gaatatggca 3960
actagattgt acacatttag aaggaaaagt tatcctggta gcagttcatg tagccagtgg 4020
atatatagaa gcagaagtta ttccagcaga aacagggcag gaaacagcat attttctttt 4080
aaaattagca ggaagatggc cagtaaaaac aatacatact gacaatggca gcaatttcac 4140
cgggtgctacg gttagggcgg cctgttgggt ggcggggaatc aagcaggaat ttggaattcc 4200
ctacaatccc caaagtcaag gtagtagata atctatgaat aaagaattaa agaaaattat 4260
aggacaggta agagatcagg ctgaacatct taagacagca gtacaaatgg cagtattcat 4320
ccacaatttt aaaagaaaaa gggggattgg ggggtacagt gcaggggaaa gaatagtaga 4380
cataatagca acagacatac aaactaaaga attacaaaaa caaattacaa aaattcaaaa 4440
ttttcggttt tattacaggg acagcagaaa tccactttgg aaaggaccag caaagctcct 4500
ctggaagggt gaaggggcag tagtaataca agataatagt gacataaaag tagtgccaag 4560
aagaaaagca aagatcatta gggattatgg aaaacagatg gcaggtgatg attgtgtggc 4620
aagtagacag gatgaggatt agaacatgga aaagttagt aaaacaccat atgtatgttt 4680
cagggaaagc taggggatgg ttttatagac atcactatga aagccctcat ccaagaataa 4740
gttcagaagt acacatccca ctaggggatg ctgattgggt aataacaaca tattgggggtc 4800
tgcatacagg agaaagagac tggcatttgg gtcagggagt ctccatagaa tggaggaaaa 4860
agagatatag cacacaagta gacctgaac tagcagacca actaattcat ctgtattact 4920
ttgactgttt ttcagactct gctataagaa aggcttatt aggacacata gttagcccta 4980
ggtgtgaata tcaagcagga cataacaagg taggatctct acaatacttg gcaactagcag 5040
cattaataac accaaaaaag ataaagccac ctttgcttag tgttaacgaa ctgacagagg 5100
atagatggaa caagccccag aagaccaagg gccacagagg gagccacaca atgaatggac 5160
actagagctt ttagaggagc ttaagaatga agctgttaga cattttcccta ggatttggct 5220
ccatggctta gggcaacata tctatgaac ttatggggat acttggggcag gagtgggaagc 5280

```

cataataaga attctgcaac aactgctgtt tatccatttt cagaattggg tgcgacata	5340
gcagaatagg cgttactcga cagaggagag caagaaatgg agccagtaga tccatagacta	5400
gagccctgga agcatccagg aagtccagcct aaaactgcct gtaccaattg ctattgtaaa	5460
aagtgttgct ttcattgccg agtttgttcc ataacaaaag ccttaggcat ctcctatggc	5520
aggaaagagc ggagacagcg acgaagagot catcagaaca gtcagactca tcaagcttct	5580
ctatcaaagc agtaagtatg acatgtaatg caacctatac caatagtagc aatagtagca	5640
ttagtagtag caataataat agcaatagtt gtgtggtcca tagtaatcat agaataatag	5700
aaaaatttaa gacaaagaaa aatagacagg ttaattgata gactaataga aagagcagaa	5760
gacagtggca atgagagtga aggagaaata tcagcacttg tggagatggg ggtggagatg	5820
gggcaccatg ctccttgagg tgttgatgat ctgtagtgot acagaaaaat tgtgggtcac	5880
agtcctattat ggggtacctg tgtggaagga agcaaccacc actctatttt gtgcactaga	5940
tgctaaagca tatgatacag aggtacataa tgtttgggcc acacatgcct gtgtaccac	6000
agaccocaac ccacaagaag tagtatttgt aaatgtgaca gaaaatttta acatgtggaa	6060
aaatgacatg gtgaacaga tgcatgagga tataatcagt ttatgggato aaagcctaaa	6120
gcoctgtgta aaattaaccc cactctgtgt tagtttaag tgcactgatt tgaagaatga	6180
tactaatacc aatagtagta gcgggagaat gataatggag aaaggagaga taaaaactg	6240
ctctttcaat atcagcaca gcataagagg taagtgacag aaagaatatg cattttttta	6300
taaacttgat ataataccaa tagataatga taactaccgc tataagtga caagttgtaa	6360
caoctcagtc attacacag cctgtccaaa ggtatccttt gagccaattc ccatacatta	6420
ttgtgccccg gctggttttg cgattctaaa atgtaataat aagacgttca atggaacagg	6480
accatgtaca aatgtcagca cagtacaatg tacacatgga attaggccag tagtatcaac	6540
tcaactgctg ttaaatggca gtctagcaga agaagaggta gtaattagat ctgtcaattt	6600
caoggacaat gctaaaccca taatagtaca gctgaacaca totgtagaaa ttaattgtac	6660
aagaccacaac aacaatacaa gaaaaagaat ccgtatccag agaggaccag ggagagcatt	6720
tgttacataa gaaaaaatag gaaatatgag acaagcaaat tgtaacatta gttaggcaaa	6780
atggaataac actttaaaac agatagctag caaattaaga gaacaatttg gaaataataa	6840
aacaataatc ttaagcaat cctcaggagg ggaccagaa attgtaacgc acagttttaa	6900
ttgtggaggg gaatttttct actgtaattc aacacaactg ttaatatgta ctgtgtttaa	6960
tagtacttgg agtactgaag ggtcaataa cactgaagga agtgacacaa tcacctccc	7020
atgcagaata aaacaaatta taaacatgtg gcagaaagta ggaaaagcaa tgtatgcccc	7080
tcccatcagt ggacaaatta gatgttcac aaatattaca gggctgctat taacaagaga	7140
tggtggtaat agcaacaatg agtccgagat cttcagacct ggaggaggag atatagggga	7200

caattggaga agtgaattat ataaatataa agtagtaaaa attgaacocat taggagtagc	7260
accocaccag gcaagagaa gagtggtaga gagagaaaa agagcagtag gaataggagc	7320
tttgttcctt ggggttcttg gagcagcagg aagcactatg ggcgcagcct caatgacgct	7380
gacggtacag gccagacaat tattgtcttg tatagtgcag cagcagaaca atttgcctgag	7440
ggctatttag gcgcaacagc atctgttgca actcacagtc tggggcatca agcagctcca	7500
ggcaagaatc ctggctgttg aaagatacct aaaggatcaa cagctccttg ggatttgggg	7560
ttgctctgga aaactcattt gccaccactgc tgtgccttg aatgctagt ggagtaataa	7620
atctctgtaa cagatttgga atcacacgac ctggatggag tgggacagag aaaltaacaa	7680
ttacacaagc ttaatacact ccttaattga agaatcgaa aaccagcaag aaaagaatga	7740
acaagaatta ttggaattag ataatgggc aagtttggg aattggttta acataacaaa	7800
ttggctgttg tatataaaat tattcataat gatagtagga ggcttgtag gtttaagaat	7860
agtttttgc gtactttcta tagtgaatag agttaggcag ggatattcac cattatcgtt	7920
tcagaaccac ctcccaacco cagggggaac cagacagccc gaaggaaatg aagaagaagg	7980
tggagagaga gacagagaca gatccattcg attagtgaac ggatccttg cacttatctg	8040
ggacgatctg cggagcctgt gctcttcag ctaccaccgc ttgagagact tactcttgat	8100
tgtaacgagg atttggaac ttctgggacg caggggggtg gaagccctca aatattggtg	8160
gaatctccta cagtattgga gtcaggaaat aaagaatagt gctgttagct tgctcaatgc	8220
cacagccata cgagtagctg aggggacaga taggggtata gaagtagtac aaggagcttg	8280
tagagctatt cgccacatac ctagaagaat aagacagggc ttggaagga ttttctata	8340
agatgggtg caagtgtgca aaaagtagtg tgattggatg gctactgta agggaaagaa	8400
tgagacgagc tgagccagca gcagataggg tgggagcagc atctcgagac ctggaaaaac	8460
atggagcaat cacaagtagc aatacagcag ctaccaatgc tgcttggtgc tggctagaag	8520
cacaagagga ggaggagggt ggttttccag tcacacctca ggtaccttta agaccaatga	8580
ottacaaggc agctgtagat cttagccact ttttaaaaga aaagggggga ctggaagggc	8640
taattcactc ccaaagaaga caagatatcc ttgatctgtg gatctaccac acacaaggct	8700
acttccctga tttagcagaac tacacaccag ggccaggggt cagatatcca ctgaccttg	8760
gatggtgcta caagctagta ccagttgagc cagataagat agaagaggcc aataaaggag	8820
agaacaccag cttgttacac cctgtgagcc tgcattggat ggatgaccog gagagagaag	8880
tgtagtagtg gaggtttgac agccgcctag catttcatca cgtggccga gagctgcac	8940
cggagtactt caagaactgc tgacatcgag cttgtctaaa gggactttcc gctggggact	9000
ttccagggag gcgtggcctg ggcgggactg gggagtggcg agccctcaga tctctcatat	9060
aagcagctgc tttttgcctg tactgggtct ctctgggttag accagatctg agcctgggag	9120

ctctctggct aactagggaa cccactgctt aagcctcaat aaagcttgcc ttgagtgctt 9180
c 9181

<210> 15
<211> 29
<212> RNA
<213> Homo sapiens

<400> 15
ggcagaucug agccuggggag cucucugcc 29

<210> 16
<211> 52
<212> RNA
<213> Homo sapiens

<400> 16
uuuuuuaggg aagaucuggc cuuccuacaa gggaaggcca gggaauuuuc uu 52

<210> 17
<211> 9413
<212> DNA
<213> Homo sapiens

<400> 17
ttgggggcga cactccacca tagatcactc ccctgtgagg aactactgtc ttcacgcaga 60
aagcgtctag ccatggcggt agtatgagtg ttgtgcagcc tccaggaccc cccctcccg 120
gagagccata gtggtctgcg gaacccgtga gtacaocgga attgccagga cgacccgggc 180
ctttcttgga tcaaccgcgt caatgcctgg agatttgggc gtgccccgcg gagactgcta 240
gccgagtagt gtggtgctgc gaaaggcctt gtggtactgc ctgtaggggt gcttgcgagt 300
gcccggggag gtctogtaga ccgtgcacaa tgagcacaaa tcctaaacct caaagaaaaa 360
ccaaacgtaa caccaaccgc gcgccacagg acgttaagtt cccggggcgt ggtcagatcg 420
ttggtggagt ttacctgttg ccgcgcaggg gccccaggtt ggggtgtgcg gcgactagga 480
agacttccga gcggtcgcaa cctcgtggaa ggcgacaacc tatcccccaag gctcgccggc 540
ccgagggtag gacctgggct cagccccgggt acccttggcc cctctatggc aacgagggta 600

tgggggtgggc aggatggctc ctgtcaccoc gtggctctcg gcctagttag ggccccacag	660
acccccggcg taggtcgcgat aatttgggta aggtcatoga tacccttaca tgcggtctcg	720
ccgacotcat ggggtacatt ccgcttgctg ggcggccctc agggggcgct gccaggggcc	780
tggcacatgg tgtccggggt ctggaggacg gogtgaacta tgcaacaggg aatctgcccg	840
gttgctcttt ctctatcttc ctcttagctt tgctgtcttg tttagaccac ccagcttccg	900
cttacagagg ggcgaacgtg tccgggatat accatgtcac gaacgactgc tccaactcaa	960
gtatttgta tgaggcagcg gacatgatca tgcacacccc cgggtgcgtg ccctgcgtcc	1020
gggagagtaa tttctccgtg tctgggtag cgctcactcc cagctcgcg gccaggaaaca	1080
gcagcatccc caccacgaca atacgacgcc acgtcgattt gctogttggg gcggtgctc	1140
tctgttccgc tatgtacgtt ggggatctct gcggatccgt ttttctctgc tccagctgt	1200
tcacottctc acctcgccgg tatgagacgg tacaagattg caattgtcca atctatccc	1260
gncacgtatc aggtcacgcg atggcttggg atatgatgat gaactggtea cctacaacgg	1320
ccctagtgtg atcgagacta ctccgagatc caaagccgt cgtggacatg gtgcccgggg	1380
cccaactggg tgtccatagc ggccttgctc actattccat ggtggggaac tgggctaagg	1440
tcttgattgt gatgctactc tttgctggcg ttgacgggca caccacgctg acagggggaa	1500
gggtagcctc cagcacccag agcctcgtgt cctggctctc acaaggccca tctcagaaaa	1560
tccaactcgt gaacaccaac ggcagctggc acatcaacag gaccgctctg aattgcaatg	1620
actccctcca aactgggttc attgctgcgc tgttctacgc acacaggctc aaocgctccg	1680
ggtgcccaga gcgcattggc agctgccgcc ccacgatga gttcgtcag ggggtgggtc	1740
ccatcactca tgatatgcct gagagctcgg accagaggcc atattgtctg cactacgcgc	1800
ctcgacgctg cgggatcgtg cctgcgtcgc aggtgtgttg tccagtgtat tgcttctc	1860
cgagccctgt ttagtggggg acgaccgctc gtttcggcgc tctcactgat agctgggggg	1920
agaatgagac agacgtgctg ctacttagca acaogcgccc gcctcaaggc aactggtttg	1980
ggtgcacgtg gatgaacagc actgggttca ccaagacgtg cgggggccct ccgtgcaaca	2040
tccggggggg cggcaacaac accctggctc gccccaagga ttgcttccgg aagcaccccc	2100
aggccactta cacaagtgtg ggctcggggc cctggttgac acccaggtgc atggttgact	2160
accoatacag gctctggcac taccctcgca ctgttaactt tacctctttt aaggtcagga	2220
tgtagtggg gggcgtggag cacaggctca atgotgcacg caattggact cgaggagagc	2280
gctgtgactt ggaggacagc gataggtcag aactcagccc gctgctgctg tctacaacag	2340
agtggcagat actgcctctg tcttccacca cctaccggc cctgtccact ggcttgatcc	2400
atcttcacag gaacatcgtg gacgtgcaat acctgtacgg tatagggtcg gcagttgtct	2460
cccttgaat caaatgggag tatatcctgt tgcttttctc tcttctggcg gacgcgcgcg	2520

tctgtgcctg cttgtggatg atgctgctga tagccaggc tgaggccacc ttagagaacc 2580
 tggtagctcct caatgoggog tctgtggccg gaggcgatgg ccttctctcc ttoctcgtgt 2640
 tcttctgcgc cgcctggtac atcaaaggca ggclgggtccc tggggcggca tatgctctct 2700
 atggcgatg ggcgttgctc ctgctcttgc tggccttacc accacgagct tatgccatgg 2760
 accgagagat ggcgtgcacg tgcggaggcg cggtttttgt aggtctggta ctcttgacct 2820
 tgtaaccata ctataagggt ttcctcgcta ggctcatatg gtggttaca tttttatca 2880
 ccagagccga ggcgacattg caagtgtggg tccccctct caatgttcgg ggaggccggg 2940
 atgccatcat cctccttaca tgcggggtcc atccagagct aatctttgac atcaccaaac 3000
 tctgtctcgc catactcggg cgcctcatgg tgctccaggc tggcataact agagtgcgt 3060
 actttgtacg cgcctcaggg ctcactcgtg catgcatggt agtgcggaag gtgcgtggag 3120
 gccactatgt ccaaatggcc ttcgatgaag tggcggcgct gacaggtagc tacgtatatg 3180
 accatcttac tccactcggg gattggggcc acgcgggcct acgagacctt gcggtggcag 3240
 tagagcccg cgtctctctc gacatggaga ctaaactcat cacctggggg gcagacaccg 3300
 cggcggtgtg ggacatcacc tcgggtctac cagtctccgc ccgaaggggg aaggagatac 3360
 ttctaggacc ggccgatagt tttggagagc aggggtggcg gctccttgcg cctatcacgg 3420
 cctattccca acaaacggcg ggccgtgctg gctgtatcat cactagccto acaggtcggg 3480
 acaagaacca ggtcgatggg gaggttcagg tgctctccac cgcaacgcaa tctttcctgg 3540
 cgacotgcgt caatggcggt tgttggaocg tctaccatgg tgocggctcg aagacctgg 3600
 cgggcccgaa ggggtccaac acccaaatgt acaccaatgt agaccaggac ctgcctggct 3660
 ggccggcgcc ccccgggggc cgcctccatg caccgtgcac ctgcggcagc tcggaccttt 3720
 acttggtcac gaggcatgct gatgtcgttc cggtgcccg cggggcgac agcaggggga 3780
 gctgctttc ccccgagccc atctcctacc tgaagggtc ctcggttgga ccactgcttt 3840
 gcccttcggg gcaogttgta ggcattctcc gggctgctgt gtgcacccgg ggggttgoga 3900
 agggggtgga ctctataccc gttgagtcta tggaaactac catgcggtct ccggtcttca 3960
 cagacaaetc atccctcccg gcogtacgcg aaacattoca agtggcacat ttacaogcto 4020
 ccactggcag cggcaagagc accaaagtgc cgggtgcata tgcagcccaa gggtaacaag 4080
 tgctcgtcct aaacccgtcc gttgcggcca cattgggctt tggagcgtat atgtccaagg 4140
 cacatggcat cgagcctaac atcagaactg gggttaaggac catcaaccag ggggccacca 4200
 tcacgtactc cactctattg aagttccttg ccgacggtg atgctcggg ggcgcctatg 4260
 acatcataat atgtgatgaa tgccactcaa ctgactcgac taccatcttg ggcacgggca 4320
 cagtctcgga tcaggcagag acggtctggg cgcggctcgt cgtgctcgcc accgccacgc 4380
 ctccgggata gatcacogtg ccacacccca acatcgagga agtgccctg tccaacactg 4440

gagagattcc cttctatggc aaagccatcc ccattgaggc catcaagggg ggaaggcacc 4500
 tcattctctg ccattccaag aagaagtgtg acgagctcgc cgc aaagctg acaggcctcg 4560
 gactcaatgc tgtagcgtat taccgggggtc tcgatgtgtc cgtcataccg actagcggag 4620
 acgtcgttgt cgtggcaaca gacgctctaa tgacgggttt taccggcgac ttgactcag 4680
 tgatcgactg caacacatgt gtcacccaga cagtcgattt cagcttggat cccaccttca 4740
 ccattgagac gacaacgctg ccccaagacg cgggtgcgcg tgcgcagcgg cgaggtagga 4800
 ctggcagggg caggagtggc atctacaggt ttgtgactcc aggagaacgg cctcaggcca 4860
 tgttcgactc ctcggtcctg tgtgagtgtc atgacgcagg ctgcgcttgg tatgagctca 4920
 cgcccgctga gacctcggtt aggttcgagg cttaacctaa tacaccaggg ttgcccgtct 4980
 gccaggacca cctagagtgc tgggagagcg tcttcacagg cctcaccacc atagatgcc 5040
 acttcttctc ccagaccaaa caggcaggag acaacctccc ctacctggta gcataccaag 5100
 ccacagtgtg cgccagggtc cagggtccac ctccatctgt ggaccaaagt tggaaagtgtc 5160
 tcatacggct aaagcccaca ctgcatgggc caacgcccct gctgtacagg ctaggagccg 5220
 ttcaaaatga ggtcactctc acacaccccc taaccaaata catcatggca tgcattgtcg 5280
 ctgacctgga ggtcgtcact agcacctggg tgctagttag cggaatcctt gcggctctgg 5340
 ccggtactg cctgacgaca ggcagcgtgg tcatttgtgg caggatcacc ttgtccggga 5400
 ggccagctgt tattcccgac aggaagatcc tctaccagga gttcgatgag atggaagagt 5460
 gtgcttcaca cctcccttac atcgagcaag gaatgcagct cgccgagcaa ttcaaacaga 5520
 aggcgctcgg attgctgcaa acagccacca agcaagcgga ggctgtgtct ccgctgtgtg 5580
 agtccaaagt gcgagccctt gaggtctctt gggcgaaaca catgtggaac ttcatcagcg 5640
 ggatacagta cttggcagcg ctatccactc tgcctggaaa ccccgcgata gcattattga 5700
 tggcttttac agcctctatc accagccgcg tcaccaccca aaataccctc ctgtttaaca 5760
 tcttgggggg atgggtggct gcccaactcg ctccccccag cgctgcttgg gctttcgtgg 5820
 gcgcggcat tgccggtgcg gccgttgcca gcatagtct cgggaaggtc cttgtggaca 5880
 ttctggcggg ctatggggcg ggggtggctg ggcactcgtt ggcotttaag gtcattgagcg 5940
 gcgagatgcc ctccactgag gatctggtta atttactccc tgccatcctt tctcctggcg 6000
 cctggttgtt cggggtcgtg tgcgcagcaa tactgctcgt gcacgtgggc ccgggagagg 6060
 gggtgtgca gtgatgaac cggctgatag cgttcgcttc gcggggtaac caagtctccc 6120
 ccacgcacta tgtgcccgag agcgacgcgg cggcgctgtt tactcagatc ctctccagcc 6180
 ttaccatcac tcagttgctg aagaggcttc atcagtggtt taatgaggac tgctccaagc 6240
 cttgttccgg ctctggtgcta aaggatgttt gggactggat atgcacgggt ttgagtgtact 6300
 tcaagacttg gctccagtc aagctcctgc cgcggttacc gggactccct ttctgtcat 6360

gccaacgcgg gtacaaggga gtctggcggg gggatggcat catgcaaacc acctgccat 6420
 gtggagcaca gatcacccga catgtcaaaa atggctccat gaggattgtt gggccaaaaa 6480
 cctgcagcaa cactggcat ggaacattcc ccatcaacgc atacaccacg ggcacctgca 6540
 cgcctcccc agcgcggaac tattccaggg cgtctggcg ggtggctgct gaggagtacg 6600
 tggaggttac cggggtggg gatttccact acgtgacggg catgaccact gacaacgtga 6660
 aatgcccatg ccagggttcca gccctgaat ttttcacgga ggtggatgga gtacggttgc 6720
 acaggtatgc tccagtgtgc aaacctctcc tacgagagga ggtcgtattc caggtcgggc 6780
 tcaaccagta cctggctcgg tcacagctcc catgtgagcc cgaaccggat gtggcagtg 6840
 tcaactccat gtcacccgac cctctcata ttacagcaga gacggccaag cgtaggctgg 6900
 ccagggggtc tccccctcc ttggccagct ctacagtag ccagttgtct gcgccttctt 6960
 tgaaggcgac atgtactacc catcatgact ccccgagcc tgacctcacc gaggccaacc 7020
 tctgtggcg gcagagagat ggcgggaaca tcaccgtgt ggagtcagaa aataagggtg 7080
 taatcctgga ctctttcgat ccgattcggg cggtaggaga tgagagggaa atatccgtcc 7140
 cggcggagat cctgcgaaa ccaggaagt tccccacg cgttgccata tggcaccgcc 7200
 cggattacaa cctccaact ctagagtctt ggaaggaccc ggaactacgt cccccgtgg 7260
 tacacgggtg ccttttgcca tctaccaagg ccccccacat accacctcca cggaggaaga 7320
 ggacggtgt cctgacagag tccaccgtgt cttctgcctt ggcggagctc gctaactaga 7380
 cctttggcag ctccgggtcg tcggccgttg acagcggcac ggcgaactgc cctcccgatc 7440
 aggcctccga cgcgggcgac aaaggatccg acgttgagtc gtactctcc atgcccccc 7500
 tcgagggaga gccaggggac ccgacctca gcgacgggtc ttggtctacc gtgacgggg 7560
 aagctgtga ggcgtctgc tgcctgctca tgcctatac atggacaggt gccttgatca 7620
 cgcctgcgc tgcggaggag agcaagttgc ccatcaatcc gttgagcaac tctttgctgc 7680
 gtcaccacag tatggtctac tcacacaacat ctgcagcgc aagtcctcgg cagaagaagg 7740
 tcaccttga cagaactcaa gtcctggacg accactacc ggaactgtc aaggagatga 7800
 aggcgaaggc gtccacagtt aaggctaggc ttctatctat agaggaggcc tgcaaacatga 7860
 cgcgccca ca ttcgccaaa tccaaatttg gctacggggc gaaggacgtc cggagcctat 7920
 ccagcagggc cgtcaaccac atccgctccg tgtgggagga cttgctggaa gacactgaaa 7980
 caccaattga taccaccatc atggcaaaaa atgaggtttt ctgcgtccaa ccagagaaa 8040
 gaggccgcaa gccagctcgc cttatcgtat tcccagacct gggggtacgt gtatgcgaga 8100
 agatggccct ttacagcgtg gtctccaccc ttctcaggc cgtgatgggc cctcctaacg 8160
 gattccagta ctctcctggg cagcgggtcg agttcctggg gaatacctlg aaatcaaaga 8220
 aatgcctat gggcttctca tatgacacc gctgctttga ctcaacggtc actgagaatg 8280

```

acatccgtac tgaggaaatca atttaccat gttgtgactt ggccccgaa gccaggcagg 8340
ccataaggtc gctcacagag cggctttatg tgggggtcc cctgactaat tcgaaggggc 8400
agaactgcgg ttatcgccgg tgcgcgcgaa gtggcgtgct gacgactagc tgcggcaaca 8460
ccctcacatg ttacttgaag gccactgcgg cctgtcgagc tgcaaagctc caggactgca 8520
cgatgctcgt gaacggagag gaccttctgt ttatctgtga gagtgcggga acccaggagg 8580
atgcggcggc cctacgagcc ttacgggagg ctatgactag gtattccgcc cccccgggg 8640
acccgcccc accagaatac gacttggagc tgataacgtc atgtccctcc aatgtgtcgg 8700
tcgcgcacga tgcattccgc aaaaggggtg actacctcac ccgtgacccc accaccccc 8760
tcgcacgggc tgcgtgggag acagttagac caactccagt caactcctgg ctaggcaata 8820
tcattcatgta tgcgcccaco ctatggggca ggatgattct gatgactcat ttcttctcta 8880
tcctttctagc tcaggagcaa cttgaaaaag ccctggattg tcagatctac ggggcctggt 8940
actocattga gccacttgac ctacctcaga tcattgaacg actccatggt cttagcgcat 9000
tttcaactca cagttactct ccaggtgaga tcaatagggt ggcttcacgc ctacggaaac 9060
ttgggttacc gcctttcgga gtcgtgagac atcggggcag aagtgtccgc gctaagctac 9120
tgtcccaggg ggggagggct gccacttgcg gcaagtaoct cttcaactgg gcagtaaga 9180
ccaagcttaa actcaactca atcccgctg cgtcccagct agacttgcgc ggctgggtcg 9240
ttgctggtta caacggggga gacatatatc acagcctgtc tcgtgccga ccccgttggt 9300
tcattgttgt cctactccta cttttcttag gggtaggcac ctacctgtcc ccaacccggt 9360
gaacggggag ctaaccactc caggccaata ggccattccc tttttttttt ttc 9413

```

<210> 18

<211> 328

<212> RNA

<213> Homo sapiens

<400> 18

```

uugggggcga cacuccacca uagaucacuc ccugugagg aaucacuguc uuacgcgaga 60
aagcgucua gccaugcggu aguaugagug uugugcagcc uccaggaccc cccucccg 120
gagagccaua guggucugcg gaacccguga guacaccgga auugccagga cgaacggguc 180
cuuucuuuga ucaaccgcgu caaugccug agauuvggc gugcccccgc gagaugcuu 240
gccgaguagu guugggucgc gaaaggccuu gugguacugc cugauagggu gcuuugcagu 300
gccccgggag gucuuguaga ccgugcau 328

```

<210> 19
 <211> 14
 <212> RNA
 <213> Homo sapiens

<400> 19
 auuugggcgu gccc 14

<210> 20
 <211> 27
 <212> RNA
 <213> Homo sapiens

<400> 20
 gccgaguagu guugggucgc gaaagcc 27

<210> 21
 <211> 340
 <212> DNA
 <213> Homo sapiens

<400> 21
 atgggaggag ggaagctcat cagtggggcc acgagctgag tgcgtcctgt cactccactc 60
 ccatgtccct tgggaaggctc tgagactagg gccagaggcg gccctaacag ggctctccct 120
 gagcttcagg gaggtgagtt cccagagAAC ggggctccgc gcgaggctcag actgggcagg 180
 agatgccgtg gaccccgccc ttcggggagg ggcccggcgg atgcctcctt tgcgggagct 240
 tggAACagac tcaogccag cgaagtgagt tcaatggctg aggtgaggta cccgcaggg 300
 gacotcataa cccaattcag accactctcc tccgccatt 340

<210> 22
 <211> 349
 <212> DNA
 <213> Homo sapiens

<400> 22

gaggaaagtc cgggctcaca cagtctgaga tgattgtagt gttcgtgctt gatgaaacaa 60
 taaatcaagg cattaatttg acggcaatga aatatcctaa gtctttcgat atggatagag 120
 taatttgaaa gtgccacagt gacgtagcctt ttatagaaat ataaaagggtg gaacgcggta 180
 aaccctctga gtgagcaatc caaatttggt aggagcactt gtttaacgga attcaacgta 240
 taaacgagac acacttcgcg aaatgaagtg gtgtagacag atgggtatca cctgagtacc 300
 agtgtgacta gtgcacgtga tgagtacgat ggaacagaac gcggcttat 349

<210> 23

<211> 377

<212> DNA

<213> Homo sapiens

<400> 23
 gaagctgacc agacagtgc cgttcgtcg tcgtcctctt cgggggagac gggcggaggg 60
 gaggaaagtc cgggctccat agggcagggt gccaggtaac gctctggggg gaaaccacg 120
 accagtgcac cagagagcaa accgccgatg gcccgcgcaa cggggatcag gtaagggtga 180
 aagggtcgcg taagagcgca ccgcgcggct ggtaacagtc cgtggcacgg taaactccac 240
 ccggagcaag gccaaatagg gggtcataag gtacggcccg tactgaaccc gggtaggctg 300
 ctttagccag tgagcgattg ctggcctaga tgaatgactg tccacgacag aaccggcctt 360
 atcggtcagt ttcacct 377

<210> 24

<211> 38110

<212> DNA

<213> Homo sapiens

<400> 24
 ccaccgggta cgatcttgcc gaccatggcc ccacaatagg gccggggaga cccggcgta 60
 gtggtgggag gcaacggtcag taacgtctgc gcaacacggg gttgactgac gggcaatata 120
 ggctccatag cgtcggccgc ggatacagta aaggagcatt ctgtgacgga aaagacgccc 180
 gacgagctct tcaaaactgc caaggacgag aaggtcgaat atgtcgacgt ccggttctgt 240
 gacctgectg gcatcatgca gcacttcacg attccggcct cggcctttga caagagcggt 300
 tttagacgag gcttggcctt tgacggctcg tcgattcgcg ggttccagtc gatccacgaa 360
 tccgacatgt tgctctctcc cgatcccgag acggcgcgca tcgacccggt ccgcgcggcc 420

aagacgtga atatcaactt ctttgtgcac gaccggttca ccttgagacc gtactccgcg 480
gaccgcgcga acatgcgcgc caaggccgag aactacctga tcagcactgg catcgccgac 540
acgcatact tcggcgccga ggcgcggttc tacattttcg attcgggtgag cttagactcg 600
cgcgccaaacg gtcctctteta cgagggtggac gccatctcgg ggtggtggaa caccggcgcg 660
gcgacgcgag ccgacggcag tcccaaccgg ggctacaagg tccgccacaa gggcgggtat 720
ttccactggt cccccaacga ccaatacgtc gacctgcgcg acaagatgct gaccaacctg 780
atcaactcgg gtttcatctt ggagaaggcg caccacgagg tgggcagcgg cgcgacggcc 840
gagatcaact accagttcaa ttgcgtgctg cgcgcgcgcg acgacatgca gttgtacaag 900
tatcatatca agaacaccgc ctggcagaac ggcaaaacgg tcacgttcat gccaaacgcg 960
ctgttcggcg acaacgggtc cgcatgcac tgcattcagt cgctgtggaa ggacggggcc 1020
ccgctgatgt acgacgagac gggttatgcc ggtctgtcgg acacggcccg tcattacatc 1080
ggcgccctgt tacaccacgc gccgtgcgtg ctggccttca ccaaccgcac ggtgaactcc 1140
tacaagcgcc tggttccggg ttaogaggcc ccgatcaacc tggttctatg ccagcgcaac 1200
cggctggcat gcgtgcgcac ccgatcacc ggcagcaacc cgaaggccaa cgcgctggag 1260
ttccgaagcc ccgactgcgc gggcaaccgg tatctggcgt tctcgccat gctgatggca 1320
ggcctggagc gttatcaaga caagatcgag ccgagggcgc ccgtcgacaa ggatctctac 1380
gagctccgc cgggaaggcg cgcgagtac ccgcagactc cgaccagctc gtcagatgtg 1440
atcgaccgtc tcgagggcca ccacgaatac ctacccgaag gaggggtgtt cacaacgac 1500
ctgatcgaga cgtggatcag tttaagcgc gaaaacgaga tcgagccggt caacatccgg 1560
ccgcatccct acgaattcgc gctgtactac gacgtttaag gactcttcgc agtcggggtg 1620
tagagggagc ggcgtgtcgt tgccaggggc ggctgcgagg tttttcgatg ggtgacggtg 1680
gccggcaacg gcgcgcgcac caccgctcgc aagagcccg ttaagaacgt tcaaggacgt 1740
ttcagccggg tgccacaacc cgtttggcaa tcattctccg accgcgcgagc gggttgtctt 1800
tcacatgcgc cgaacctcaa gccacgtcgt cgcacggcg tgcgtgcgcg gccggttcag 1860
gttaagtgtc ggggattcgt cgtgcggcg gcgctccacg ctgaccaacg ggcagtcgaa 1920
ctccgaaca cttttgcgac tacgccttt gccgcgcgcg tcaccgtag gtatgtgtcc 1980
aggaattccc caccgtgcgc gtttcgccag ccggccgcga ccgcgaccgc attgagctgg 2040
cgcccggtc ccggcagctg gtcggtgggc ttgcgcgcga ccaacaccag cgcgttgcg 2100
gcccggttg ccgtcagcca gccctgacgc agcagctcca cgtcggctgc ggaaccaga 2160
tcggcgccgc cgatgacatc cagggtatgc agcgtcgagg tgttgtcgag ggcgggaacc 2220
tggtgcgcac gctgtagctg cagcaactgc acggtccatt cgatgtcgcg cagtcgcgcg 2280
cggcccgatt tgggtgtgtg gttggggctg gcacgcgcgc gcaaccgcgc ggactcgata 2340

cgggccttga	tgccgcgaat	ctcgcgcacc	gagtcagcgg	acacaccgcg	gggcggtatac	2400
cgcgttttgt	cgaccatccg	taggaatcgc	tgacccaact	cgcatcgcgc	ggcaaccgcg	2460
tgtgcgcgta	gcagggcctg	gatctcccat	ggctgtgccc	actgctcgta	gtatgcggcg	2520
taggacccca	gggtgcggac	cagcggaccg	ttgcggccct	cgggtcgcaa	attggcgctg	2580
agctccagcg	gcgtagcgac	gctgggtgtc	cccagcagcg	ccgaaccgcg	ctcggcgatc	2640
gatgtcgacc	atttcaccgc	ccgtgcacgc	tcgacgcggc	tgcccggtgc	acagacgaac	2700
atcacgtcgg	catccgaacc	gtagcccaac	tcggcaccac	ccagccgacc	catgccgatg	2760
accgcgatgg	cgcgcggggc	gcgatcgtcg	tcgggaaggg	tgccccggat	catgacgtcc	2820
agcgcggcct	gcagcaccgc	cacccacacc	gacgtcaacg	cccgccacac	ctcggtgacc	2880
tcgagcaggc	cgagcaggct	cgcggaaccg	atgcgggcca	gctctcgacg	acgcagcgtg	2940
cgcgcgcggg	cgatggcccg	ctccgggtcg	gggtagcggc	tcgccgaggg	gatcagcgcc	3000
cgagccacgg	cgcgcgggtc	ggtctcgagc	agcttcgggc	ccgcaggccc	gtctctgtac	3060
tgtgtgatga	ccgcggcgcg	gcgcataaac	agatccggca	catacgccga	ggtaaccaag	3120
acatgcgatg	gcgccttggc	caccccgggc	ttgtcccgca	gcgtggccag	gtaccagctt	3180
tcggtggcca	gcgcctcaat	gagccgcggg	taggccagca	gtccgcgctc	gggatcgggg	3240
gcatacgaca	tcaggtccag	cagcctgggc	agcagcaccg	actgcacccg	tcgcgcgcgg	3300
ccgctttgat	tgaccaaacc	cgacatgtgt	ttcaacgcgg	tcgcgggtcc	ctcgtagccc	3360
agcgcggcca	gcgcggcgcc	cgccgctccc	aacgtcatgc	cgtggcgcat	ctccaaaccg	3420
gtcggggcca	tcgattccag	cagcgggtga	tagaagagtt	tggtgtgtaa	cttcgacacc	3480
cgcacgttct	gctctcttag	ttctctccgc	agcaccocgg	ccgcacgttt	tcggccatcg	3540
ggcgggatgt	gggcccgcgg	cgcagccagc	cgcactgcct	cctcgtcttc	gggatcggga	3600
agcaggtggg	tgcgcttgag	ccgctgcaac	tgcatgctgt	gctcgagcag	cctgaggaac	3660
tcatacgacg	cggctcatgt	cgcgcgctcc	tcacgcgcga	tgtagccgcc	ttgcgccaac	3720
gccgccaatg	cgctccaccg	ggacgcacc	cgtaacgact	cgtcgctacg	ggcatgaacc	3780
agctgcagta	gctgtacggc	gaactccacg	tcgcgcaatc	cgcgcgtgcc	gagtttgagc	3840
tcgcggccgc	ggacatcggc	gggcaccagc	tgctccaccc	gcgcgcgcgc	ggctcgcaac	3900
tcgaccacaa	agttcttcgc	ctcgcaggct	cgcacaccca	tcggcatcaa	ggcggtcagg	3960
taacgcctgc	caagtctccg	gtcgccaacg	actggcgcgt	ctttagcaaa	cgcctgaaac	4020
tcccaggtct	tgcccacggc	ctggtagtag	gcgatgtgcg	acccgagcgt	acggaccagc	4080
tcccctgttc	gcctctccgg	acgcaggggc	gcgtccacct	cgaaaaaggg	cgcgcaggcc	4140
accgcgatca	tctcgtcggc	cacgcgcgcg	ttgcgcgggt	cggagcgcgc	ggcaacgaat	4200
atgacatcga	cgtcgtcgac	gtagttcagt	tcgcgcgcac	cgcacttgcc	catcgcgatg	4260

accgccaggc gcgggtggcgg gtgctcgccg cacacgctcg cctcggccac gcgcagcgcc 4320
 gccgccagag cggcgctcgcg ggcgtcgccg aggcgtcgcg ccaccacggg gaatggcagc 4380
 accggttcgt cctcgaccgt cgcggccagg tcgagagcgg ccagcattag caogtagtcg 4440
 cggtaactggg ttgcgaatcg gtgcacgagc gagcccgcca taccctccga ttctcgcagc 4500
 cactcgacga acgaccgctg cagctggcca tgggacggca gtgtgacctt gccccgcagc 4560
 aatttcacgg actgcggatg ggcgaccagg tgatcgccca acgbcagcga cgagcccgagc 4620
 accgagaaca gccgcccgcg cagactcgct tcgcgacgca gagcccggtt gagctcgtec 4680
 catccggtgt ctggattctc cgacagccgg atcaaggcgc gcagcgcggc atcgcggtcc 4740
 ggagcgcggt acagcgacca cagcaggtcg acgtgcgcct gatcctcgtg ccgatccac 4800
 cccagctgag ccagacgctc accagcaggg gggtcaacta atccgagcgc gccaacgctg 4860
 ggcaacttcg gccgctcgct ggcgagtttg gtcacgacca cgacggtagc gcaaagcgcg 4920
 tcggcgctcg atcaaacggg agatctgggc tacagcgaca ggtaggtagc cagctcgatat 4980
 ggcgtagcgt ggcgtcggta gttcgccac tcogtgcgct tgttgcgcaa gaaaagtca 5040
 aaaaactgct cccccaaggc ctccgcgacg agttcggagg cctccatggc gcgcagcgca 5100
 ctatccaaac tggcagggca ttctcggtac cccatcgctc ggcgttccto ggggtgtagg 5160
 tcccatcgtg tgtctcggc ctgcggggcc agcacgtaac ccttctctac accccgcaat 5220
 ccccgcgcca gcagcacggc gaatgtcaga tagggattgc acgcccgaatc agggctgogt 5280
 acttcgaccc gccgcgacga ggtcttgtgc ggcgtgtaca tcggcacccg cactaggggc 5340
 gatcggttgg cggcccccga cgacgcggcc gtgggcgctt cgcgcgctg caccagcgcc 5400
 ttgtaagagt tgacccactg atttgtgacc gcgctgatct cgcaagcgtg ctccaggatc 5460
 ccggcgatga acgatttacc caettccgac agctgcagcg gatcatcagc gctgtggaac 5520
 gcgttgacat caccctcgaa caggctcatg tgggtgtgca tcgccgagcc cgggtgctgg 5580
 ccgaatggct tgggcatgaa cgacgcccgg gcgcctctt ccagcgagc ttctttgatg 5640
 acgtagcggg aggtcatcac gttgtcagc atcgacagag cgtcgcaaaa ccgcaggtcg 5700
 atctctgct gcccggtg ccttcgtga tggctgaact ccaccgagat gcccatgaat 5760
 tcaggggcat cgatcgcggt gcggcgaaa ttcaaggcgg agtcgtgcac cgcttggtcg 5820
 aaatagccgg cgttgtgcac cgggacgggc accgaccgct cctcggtgtcc gggcttgagc 5880
 aggaagaact cgatttcggg atgcacgtag caggagaagc cgagtctgac ggccctcgtc 5940
 agctgcgcc gccaacacgt ccgcgggtcc gccacgacg gcgagcgctc cggcatggtg 6000
 atgtcgcaaa acatccgcgc tgagtgtgtg tggccggaac tggtggccca gggcagcacc 6060
 tggaaggctg acgggtccgg gtgcgccacc gtatcggtat ccgagacccg cgcaagccc 6120
 tcgatcgagg atcgtcgaa gccgatgcct tctcgaagg cgcctcgag ttcggtcggg 6180

gcgatggcga cgcacttgag gaaacgcgagc acgtctgtga accacagccg gacgaagcgg 6240
 atgtgcggtt cttccagggg acgaagaacg aattccttct gtccgtccat accctogaaca 6300
 gtatgcactg tctgttaaaa ccgtgttacc gatgcccgcc cagaagcggt gcggggcgcc 6360
 ccgcaagggg agtgcgcggt gagttcaggg cgcgcaccgc agactcgtcg gcggcaaggt 6420
 cccgtcgaga aatagtgtga tcaccgcaga gtccacacac tggttgccat cgaacaccgc 6480
 agtgtgttgg gtgcgcgcga aggtgatcag cgggtgcgcc agctggcggg ccaggtctac 6540
 ccggactga tacggagtgg ccgggtcgtg ggtggtggac accacgacga ccttgccagc 6600
 ccggcgccgc gcgcgggggt gcggcgtcga cgttgccggc accggccaca gcgcgcacag 6660
 atcgcggggg gcggatccgg tgaactgccg gttagctaagg aacggggcga cctgacggat 6720
 ccgttggtcg gcggccaccc aggcgcgtgg atcggccggt gtgggcgcat cgacgcaccg 6780
 gaccgcgttg aacgcgtcct ggtcgttgcg gtatgcgccg tctgcacccc ggccgtcata 6840
 gtcgtcggca agcaccagca agtcgcgcgc gtgcgtcccg cgtgcagcc cagcagacc 6900
 actggtcagg tacttcacgc gctgagggct gtacagcgcg ttgatggtgc ccgtcgtcgc 6960
 gtcggcgtag ctcaggccac gtggatccga cgtcttacc ggcttctgca ccagcgggtc 7020
 aaccagggcg ttgtagcggg tgaccocact ggccgagtcg gtgccagag ggacggcccg 7080
 cgagcggggc cagtgcggcg cgtagtcatt gaaagcggtc tgaatcccg ccatttggt 7140
 gatgctttcc tcgattgggc taaagcgtgg atcgatagcg ccgtcgagga ccacgcgccg 7200
 cacatgagta ccgaaccggt ccaggtaagc ggtgcccaac tcggtgccgt agctgtatcc 7260
 gaggtagtgg atctgatcgt caccctaacc ttggcgcaac atgtccatgt ccgtgcgac 7320
 ggacgcggta ccgatattgg ccaagaagct gaagcccatc cggtaacac agtctgggc 7380
 caactgccgg tagacctgt cgacgtgggt gacaccggcc ggactgtagt cggccatcg 7440
 atcgcgcggg tacgcgtcga actcgcgtgc ggtgcgacac cgcaacgcag ggttcgagtg 7500
 gcggaaccct ctcgggtcga agccaccag gtogaagtgg cggagaatgt cgggtcggc 7560
 gatcgcgggt gccatagcgg cgaacctgtc gaacgcgcac gccccgggtc cccacggatt 7620
 gaccagcagt gtcocgaato gctgtcccg cgcggggacg cggatcacg ccaacttcgc 7680
 ttgtgtccca ccgggttggt cgtagtcgac ggggacggac accgtgcgcg agcgtgcagt 7740
 gcgaatttgg cttggttcgg cgtgaaactc gcggcagctg ttccaactct gttgcggcgc 7800
 cagcaccgcg gcaaccgggg tttggcggcg gccgggttct tcagtgcgcg cggccaaccg 7860
 gggcgctgct aggggcagtc cgcgcagcag caaccggaag gacagcagcg ccgagctcaa 7920
 cggctcgcgg gccacatggt ccgcacatcgt ctcaccggcg aatacctgtg acggcgcgaa 7980
 atgatcacac cttcgtttct tcgccccgct agcacttggt gccgtggggc ggcgtggtgc 8040
 ccgcgattaa ataccgcgtc acgtactcgt caatgcagct gtcgcctcgg aataccaccg 8100

tgtgctgggt tccgtcgaag gtcagcaacg aaccgcgaag ctggttcgac aggtcgacc 8160
 cggccttcta cggcgtcgcc gggcatggg tggtgatac caccaccgac ggcactaggc 8220
 cggggcgcga gacggcatgg ggcagacttg tgggtggcac cggccagaac gcgcaggtgc 8280
 ccagcggcgc atcacgggtg aacttccgac agctcatgaa cggtcgacgc tcccgggcgc 8340
 ggcggctctc gtcatgacc ttgtcgcgat cggtaaccgg gggctgatcg acgaattga 8400
 tcgccaccgc cgcgtcacgc gaattgttgt agcggcgtg cagatccga cgcattgaca 8460
 tgtcgccag agccagcagg gtgtctccgc gattgtcgac cagctccgac agcccgctcg 8520
 tcaagtgttg ccacagatcc ggtgagtaca gcgccataat ggtgccccag atggcgctgc 8580
 tataactcag ccgcgcgga tccttcgtgc gcgcggcct gctgacctc gggttgtccg 8640
 ggtgaccaa cggatcgacc aggtctggt agacctcgac ggtttggcc ggtcgcgcc 8700
 ccagcgggca gccgcgcttc ttggcgagc cggcggcata gttgttgaac gcgtctgga 8760
 agcccttgcc ctggcgagc tcgccctcga tgggatcgcc attggggctc accgcaaccgt 8820
 cgagaatcat tgcgccacc cgtcgcgga atctctcgcc atacgcggag ccgattccgg 8880
 tgcggtacga gtacccagg taggtcagct tctcgtcgcc caacgcgcgc cgaatggcat 8940
 ccaggtcctt ggcgaacttg accgtccga catgggccag aaagtctctg ccatcttgt 9000
 ccacacagc accgacgaat tgetttggtc cgttctgat gtgcgccaca cctcccgcc 9060
 tgtagtcaac ctgcggctcg gccgcagcc ggtcgttgtc ggcatcgag ttgcaccaga 9120
 tcgocggcgc ggacgacgc accccggcg ggtcgaacc aaccaggctg aaccttctgt 9180
 gcaccgcct cggcaatgct tggaaagcgc ccaaggcgcc ctcgatacgc gattcgccg 9240
 gtccaccggg atttatgacc agcgaaccga tcttgtctcc cgtcgccga aagcgaatca 9300
 gcgccagcgc cgcacgctca ccatcggggc ggtcgtagtc gaccgttaca gcgagcttgc 9360
 cgataaacgc gccgcgggg atctttactt cggggtttga cgaccggcac ggtgtccact 9420
 ccacggcgtg gccagcttc ggcctccgca tacgagcgcg tcccccagcc acgggatgc 9480
 agccacaaag aaccaagcc accggcgcca gcgcggcca gatcaacagc atgcgcgcga 9540
 tcttctgcgc ggcagacgc ctcatgccca caatgctgcc agagcagacc cgagatcctg 9600
 gccagggcc accgtcgcc gactaacgg ccgctgccag cagtctgcc atcgccgatg 9660
 gcgaactcgt cggccatccc ccatacgtcc ggtaacagat ccgggcaaga caccgaccgc 9720
 tcgaccggat ccggcacggc cgcgtcgcc tcggcggtgc acaactgcga catcaggttg 9780
 gcgctggcac cccgtccag ccggcatggt gcaccttgcc catcgccga gggcgatccc 9840
 cgatgcgctc cacccttgc acgaaccat ctcccacggc ggtcgccgc agcagcgcga 9900
 tgtggcgcca gatctccgag agttcgccc gcccgcccg cgacggcaac ccgatgcgt 9960
 gcaagtgaac atcgatgta ggttcaaggt tcagcgact gctggcaagc ttttccgaa 10020

accgcggcct cgccttgatc tggagtcaga acgcgtcacg cagccggtea aaggcgtaac 10080
 ccattgctcga gcaaacatgc atgggctgag tggacgttcc cagacacagc aactggcgctc 10140
 caggccaactg agccgctgca tgcgcgatgg tatgccgatg ggggccccgg gcgcgtctga 10200
 ggggaagaag tggcagactg tcagggtccg acgaaccggg ggaccctaac gggccacgag 10260
 gatcgaccgg accaccatta gggacagtga tgtctgagca gactatctat ggggccaata 10320
 ccccgaggag ctccggggcg cggaccaaga tccgcaccca ccacctacag agatggaag 10380
 ccgacggcca caagtgggcc atgctgacgg cctacgacta ttgacgggcc cggatcttgc 10440
 acgagggcgg catcccggtg ctgctggtcg gtgattcggc ggccaacgtc gtgtacggct 10500
 acgacaccac cgtgccgatc tccatcgacg agctgatccc gctggtccgt ggcgtggtgc 10560
 ggggtgcccc gcacgcaactg gtcgtcgccg acctgccgtt cggcagctac gaggcggggc 10620
 ccaccgcgcg gttggccgcc gccaccgggt tctcaaggga cggcgccgca catcggtga 10680
 agctcgaggg cgggtgagcg gtggccgagc aaatcgctcg tctgaccgcg gcgggcaatc 10740
 cgggtgatggc acacatcgcc ttacccccg aaagcgtcaa cacctggggc ggtctccggg 10800
 tgcaggcgcc cggcgacgcc gccgaacaaa ccacgcgga cgcgatcgcc gtcgccgaag 10860
 ccggagcgtt tgccgtctgt atggagatgg tgcccgccga gttggccacc cagatcaccg 10920
 gcaagcttac cattccgacg gtcgggatcg gcgtggggc caactcgac gccaggtcc 10980
 tggatggcca ggacatggcc gggttcagcg gcgccaagac cgcccgcttc gtcaaacggt 11040
 atgccgatgt cgggtgtgaa ctacgcgctg ctgcaatgca atacgccaag gaggtggccg 11100
 gcggggtatt ccccgctgac gaacacagtt tctgaccaag ccgaatcagc ccgatgcgcg 11160
 ggcatcgcg tggcgccctg gatgccgtcg acgcgggatt gccggcgcg acgcgccagc 11220
 gggaccacac cgcgtcgctg tcgccggttg agccgggggt gagccagac attcgatgtg 11280
 cccaacacca tccgccacag ccaattgat gtggcactct atgcatgcct atccccgacc 11340
 aaccaccacc gcggcgacgc atcatgacgg gaggcgaaga tgccagtaga ggcgccaga 11400
 ccagcgcgcc atctggagggt cgagcgcaag ttgacgtga tcgagtgac ggtgtcgccg 11460
 tegtctgagg gcatcgccgc ggtggttcgc gtcgagcagt cggcgaccca gcagctcgac 11520
 ggggtgtact tcgacacacc gtcgcacgac ctggcgcgca accagatcac ctgctggcgc 11580
 cgcaccggcg gcgcgcagcg cggctggcat ctgaagctgc cggccggacc gcacaagcgc 11640
 accgagatgc gagcaccgct gtcgcacatc ggcgacgctg tgccggccga gttgttggtg 11700
 gtggtgctgg cgatcgtccg cgaccagcgg gttcagcgg tcgcgcggat cagcactcac 11760
 cgcgaagacc agatcctgta cggcgccggg ggcgacgcgc tggcggaatt ctgcaacgac 11820
 gacgtcacgg catggtcgcc cggggcattc cagcgccgtg gtgcagcgga caacggccct 11880
 gccgaacagc agtggcgcga atgggaactg gaactggtca ccacggtatg gaccgccgat 11940

accaagctac tggaccggct agccaacggg ctgctcgatg ccggtgccgc acctgccgcg 12000
 caaggctcca aactggcgcg ggtgctcggt gcgacctctc ccggtgagct gcccaacggc 12060
 ccgcagccgc cggcggatcc agtacaccgc gcggtgtccg agcaagtcca gcagctgctg 12120
 ctgtgggata gggccgtgcg ggcgcagccc tatgacggcg tgcaccagat gcgagtgcgc 12180
 acccgcaaga tccgcagctt gctgacggat tccagaggat cgtttggcct gaaggaaagt 12240
 gcgtgggtca tcgatgaact gcgtgagctg gccgatgtcc tgggcgtagc ccgggacggc 12300
 gagggtactcg gtgaccgcta ccagcgcgaa ctggacggcg tggcgccgga gctggtacgc 12360
 ggccgggtgc gcgagcgctt ggtagacggg gcgcggcggc gataccagac cgggctcgcg 12420
 cgataactga tcgcattgcg gtgcgacggg tacttccgtc tgcttcgacg cctagacggc 12480
 cttgtgtccg aacgcgcccc tgccacttot ggggaggaat cggcaccggg aaccatcgat 12540
 gcggcctacc ggcgagtcgg caaagccgca aaagccgcaa agaccgccgg cgaccaggcg 12600
 ggcgaccacc accgcgacga ggcaattgcac ctgatccgca agcgcgcgaa gcgattacgc 12660
 tacaccgcgg cggctactgg ggcggacaat gtgtcacaag aagccaaagt catccagacg 12720
 ttgctaggcg atcatcaaga cagcgtggtc agccgggaac atctgatcca gcaggccata 12780
 gcgcggaaca ccgcggcgga ggacaccttc acctacggtc tgctctacca acaggaagcc 12840
 gacttgccg agcgcgtccg ggagcagctt gaagccggcg tgcgcaaact cgcacaaggc 12900
 gtccgcaag caggggattg agcccgccag ggcgggacga gttggcctgt aagccggatt 12960
 ctgttccgg ccgcaccagc caagctaaag gcggcacggc ggcgaccatc catctggaca 13020
 cacogttacc gggcgctcgg agcggcctac ccgcaggctc gggcgagcaa cctcaaggcg 13080
 cctgcgcggc cgcacttttc gtgcggcctt cttggccttg cttcgggtgg ggtttgccta 13140
 gccaccccg tcaccggaa tgctgggtcg ctcttacgc accgtttcac cttgccacc 13200
 acgaggatgg cggctctgtt tctgtggcac ttcccgga gtcacctgg attgcggtta 13260
 gcaatcaccc tgctctgtga agtcggact ttctcgact cgacgtgaa cctcgtgaat 13320
 ccacacaagc ctaacggcag ccgcggccgc ccagccaact catccgcgac gaccacgata 13380
 ccccgctggg cgggtgtcgg gccagtgtga ccgctggacg acacggctag tcggacagcc 13440
 gatccggcgg cgactcotta tcgtggactg gtgacaagg gggacaaacg cgtcgaactc 13500
 ggcgactggg acgccatcgc tgcgaggtc agcgagtacg gtggcgact gctacctcgg 13560
 ctgatcaccc ccggcgaggc cgcocggctg cgcaagctgt acgcgcagca cggcctgttt 13620
 cgctcgacgg tcgatatggc atccaagcgg tacggcgccg ggcagtatcg atatttccat 13680
 gccccctatc ccgagtgtac gagcgtctca agcaggcgct gtatcccaaa ctgctgcga 13740
 tagcgcgcaa ctgggtgggc aaactgggcc gggaggcgcc ctggccagac agccttgatg 13800
 actggttggc gagctgtcat gccgcggccc aaacccgatc cacagcgtg atgttgaagt 13860

acggcaccaa cgactggaac gccctacacc aggatctcta cggcgagttg gtgtttccgc 13920
 tgcaggtggg gatcaacctg agcgatccgg aaaccgacta caccggcgcc gagttcctgc 13980
 ttgtlgaaca gggcgctcgc gcccaatccc ggggtaccgc aatgcaactt ccgcaggggac 14040
 atggttatgt gttcacgacc cgtgatccgc cgggtcggac tagccgtggc tggtcggcat 14100
 ctccagtcgc ccatgggctt togactattc gttccggcga acgctatgcc atggggctga 14160
 tctttcacga cgcagcctga ttgcacgcca totatagata gcctgtctga ttcaccaatc 14220
 gcaccgacga tggcccatcg cgttagaact cggcgatgct cagcgatgcc agatcaagat 14280
 gcaaccgata taggacgccc gaccggcat ccaacgccag ccgaacaac attttgatgc 14340
 gcgtgacatg tgacaccacc agcaccgctg cgccttcgta gccaacgatg atccgatcac 14400
 gtcccccgcc aacccgccgc agcacgtcgt cgaagcttcc ccaccgccgg ggcgtgatgc 14460
 tgggtgctcg cagccagcga cgggtgcagct cgggatccgc ttctgcggcc tccgcgaacg 14520
 tcagccctc ccaggcgccg aagtcgggtc cgaccaggtc gtcacgacg accacgtcca 14580
 gggccagggc tctggcgccg gtcaccgcgg tgcgtaagc ccgctgtagc ggcgaggaga 14640
 ccaccgcagc gatcccgccg cgcgcgcgca gataccggc cgcgcgacca acctggcgcc 14700
 accccacctc gttcaacccc gggttgcgcg gcccgaata gggcggttgc tccgacagct 14760
 ccgtctgcc gtggcgcaac aaaagtagtc ggggtgggtg accgcggggc ccggtccagc 14820
 cgggagatgt cggtgactcg gtcgcaacga ttttggcagg atccgcatcc gcgcgagccg 14880
 attgcgcggc ggcgtccatc cgttcattgg ccaaccggtc tgcatacgtg ttcggggcac 14940
 gcggaaccca ctctagattg atctctcgaa actgggacgc caacgcctga gcctggacat 15000
 agagottcag cagatccggg tgcttgacct tccaccgccc ggacatctgc tccaccacca 15060
 gcttgagatc catcagcacc cggcgctcgg tggcacctag ttccacggcg tcgtccaaac 15120
 cggctatcag gccgcggtat tcggcgacgt tgctcgtcgc ccggcgatc gcctgcttgg 15180
 actcggccag caccggtggag tgatcggcgg tccacaccac cgcgcgctat ccggccggtc 15240
 cgggattgcc ccgcgatccg ccgtcggctt cgatgacaac tttaactcct caaatccttc 15300
 gagccgcaac aagatcgcto cgcattccgg gcagcgacac acttcatcct cggcgggcgc 15360
 cgagatctgg gccagctcgc cgcggccgat ctgatccgg caggcacccac atcgatgacc 15420
 ttgcaaccgc ccggccctcg gcccgctcc ggcccgctgt ctttcgtaga gccccgcaag 15480
 ctcggtatca agtgctccg tcagcatgtc gogttgogat gaatgttggt ggcgggcttg 15540
 gtcgatttcg gcaagtgcct cgtccaaagc ctgctggcg cgggccaggt cggcccgcaa 15600
 cgtttggagc gcccgcgact cggcggtctg ttgagcctgc agctcctcgc ggcgttccag 15660
 caccctcagc agggcatctt ccaactggc ttgacggcgt tgcaagctgt cgagctcgtg 15720
 ctgcagatca gccaatgctt tggcgtccgt tgcaccgaa gtgagcaacg accggtcccg 15780

gtgcgcacgc ttacgcaccg catgatctc cgactcaaaa cgcgacacct ggccgtccaa 15840
 gtccctccgc ggcattcgca gggccgccat cctgtcgttg gcggcggtgt gctcggcctg 15900
 cacctgctgg taagccgccg gctgcggcag atgggtagcc cgatgcgcga tccgggtcag 15960
 ctcagcatcc agcttcgccca attccagtag cgaccgttgc tgtgccactc cggctttcat 16020
 gcctgatctc tcccagtttc gtgatcgagg ttccacgggt cgggtgcgat ggtgcacaca 16080
 cgcaccggca gcgacgcgcg gaaatgagac cgcaacactt cggcgccctg gccgcaccac 16140
 gggaattcgc ttgcccaatg cgcgacgtcg atcagggccca cttgcgaagc tcggcaatgc 16200
 tcgtcggctg gatgatgtcg cagatcgccc gtaacgtacg cttgcacgtc cgcggcgccc 16260
 acggtggcaa gcaacgagtc cccggcgccg ccgcagaccg gcaccgcga caccagcagg 16320
 tcgggatccc cggcgcgccg cacaccgttc gcagtcggcg gcaacgcggc ctcacagcgg 16380
 gcaacaaagg tgcgcagcgg ttccgggtttt ggcagtcctg caatccgcc taaccgcgtg 16440
 ccgaccggcg gtggtaccag cgcgaagatg tcgaatgccg gctcctcgta aggggtgcgcg 16500
 gcgcgcacgc ccgccaacac ctcggcgccg gctcgtcgcg gtgcgacgac ctcgaccgcg 16560
 tcctcggcca cccgttcgac ggtaccgacg ctgcctatgg cgggcgacgc cccgtcgtgc 16620
 gccaggaact gcccggtacc cgcgacactc cagctgcagt gcgagtagtc gccgatatgg 16680
 ccggcacccg cctcaaagac cgtgcgccg accgcctctg agttctcgcg cggcacatag 16740
 atgaccactc tctcgagatc ggccgctcgc gccaccgggt cgagaacggc gtcgacggtc 16800
 agaccaacag cgtgtgccag cgcgtcggac acaccggcg acgcagatc ggcgttggtg 16860
 tgcgcggtaa acaacgagcg accggtccgc atcaggcggt gcaccagcac accctttggc 16920
 gtgttgccgc cgaccgtatc gacccaacgc agtaacaacg ggtggtgcac caatagcagt 16980
 ccggcctggg gaacctggtc caccaccgcc ggctgcgct ccaccgcaac ggtcaccgaa 17040
 tccaccacgt cgtcggggtc gccgcacacc agaccaccg aatccacga cttgggcaagc 17100
 cgcgcggggt aggcctggtc cagcacgtcg atgacatcgc ccagccgcac actcatcgcc 17160
 gtccctccac ctttgccca ctcggcatcg ccgccaccag caggggccac tccgggcgca 17220
 ccgcgcgccg caggtaaccg cgtccaggc cgacgaaggt gtcaccgcgg cgcaccgcga 17280
 ttcccttgct ctgcaaatag ttctgtaata cgtcagcatc ggcgatgttg aacagtagca 17340
 aaggggccgc accatcgacc acctcggcac ccaccgatct cagtcgggcc accatctccg 17400
 cgcgcagcgc cgtcaaccgc accgcacatc ctcgggcagc ggcgaccgcc cggggggcgc 17460
 agcaagcagc gatggccgtc agttgcaatg ttcccaacgg ccagtgcgct cgtgcacgg 17520
 tcaaccgagc cagcactctt ggcgagccga gcgcgtagcc caccgcgaat ccggccagcg 17580
 accacgtttt cgtcaagcta cggagcacca gcacatcggg cagcgagtca tcggccaacg 17640
 attggcgctc gccgggaacc caatcagcga acgcctcgtc gaccaccagg atgcgtcccg 17700

gccggcgtaa ctcgagcagc tgcctgcgga ggtgcagcac cgaggtgggg ttggtcggat 17760
 taccocagac gacaaggtcg gcgtgctcag gcacgtgcgc ggtgtccagc acgaacggcg 17820
 gcttttagac aacatggtgc gccgtgattc cggcagcgct caaggctatg gccggtcgg 17880
 tgaacgcggg cagcagcatt gctgcccgca ccggacttag gttgtgcagc aatgcgaatc 17940
 cctccgcgcg ccgcagcagc gggagcactt cgtcacgggt tctgccatga cgttcagcga 18000
 ccggtcttgc ccgccggtgc acatcgtcgc tgcctcgata gcgggccagc tccggcagca 18060
 gcgcggcgag ctgcgggacc aaccattccg ggggccggtc atggcggagc ttgacggcga 18120
 agtccagcac gccggggcgc acatcctgat caccgtggta gcgcgccgcg gcaacggggc 18180
 tagtgtctag actccccaca gcgtcaaaca gtagtgggcc ggtgtcggg ccaagaatcc 18240
 agagaccgcg cgacgcgttg totacgcggc gacaaccgcg acatcacagg cagctaacag 18300
 ggcgctggcg gtgatgatcg tcaggccaag cagctgtgac tgggcgatga gcacacggtc 18360
 gaatggatgt cgatggtgat ccggaagctc tcggtgcgcg agtgtgtgcg tggtaacctg 18420
 acagcggcga cgtgcccgag cggcgcatto gatcgggcac gtaagaagcc gatggctcgg 18480
 gcggcgggag cttgccgagg cggtagttga tcgcgatctc ccaggcaactg gcggccgaca 18540
 agagaatgct gttgcggagc tctgaacaa tcgcccggtg ttctgtgacg gcactccgag 18600
 ccaaactggt gttgtgatga ggtagcgctt caccgggtga agcgttcgag cacgtcgtct 18660
 gacaacggag cgtccaaatc gtcgggcacg cggtaacgcg catggtcaat gcctaaccgc 18720
 cgagtctcat gaggatcgag cggcacaagc ttgtctaccg gctcgcgcgc gcggccaatc 18780
 tcaacctctg ccgcgcgtag acgagccgca gcagctcgga caggcgtgtc ttgcctcgt 18840
 gaacccgac ccgcttcgca ggcgccaga ctttcgcgtc gaccacctgc tcacaaact 18900
 tcgcgatcat cgcctgatac cacagcgcca acgggtagcg gttgtccaa ccgcttcgtc 18960
 aacgacaatg ggatcgtgac cgacacgacc gcgagcggga ccaattgcc gcctcctcca 19020
 cgcgcgcgcg caccggcgcg atcgtcgccg ggtgaatgc cgcagctggt gatcttcgat 19080
 ctggacggca cgtgaccga ctcggcgcg ggaatcgat ccagcttcgc acacgcgtc 19140
 aaccacatcg gtgccccagt acccgaagcg gacctggcca ctacatcgt cggcccgccc 19200
 atgcatgaga cgtcgcgcgc catgggggtc ggcgaatccg ccgaggagcg gatcgtagcc 19260
 tacggggcgc actacagcgc ccgcggttgc gcgatgaaca gcttgttcga cgggatcggg 19320
 ccgctgctgg ccgaactcgc caccgcgggt gtcgggtcgg ccgtcgccac ctccaaaggca 19380
 gagccgacgc caccggcaat cctgcgccac ttcggaattg agcagcactt cgaggtcatc 19440
 gcgggcgcga gcaaccgatg ctcgcgagcg agcaaggtcg acgtgctggc ccacgcgctc 19500
 gcgcagctgc ggcgcctacc cgagcgggtg gtgatggtcg gcgaccgcag ccacgacgtc 19560
 gacggggcgg ccgcgcacgc catcgacacg gtggtggtcg gctggggcta cgggcgcgc 19620

gactttatcg acaagacctc caccaccgtc gtgacgcatg ccgcacacgat tgacgagctg 19680
 agggaggcgc taggtgtctg atccgctgca cgtcacattc gtttgtagcg gcaacatctg 19740
 ccggtcgcca atggccgaga agatgttcgc ccaacagctt cgccaccgtg gacctgggtga 19800
 cgcggtgcga gtgaccagtg cgggcaccgg gaactggcat gtaggcagtt gcgcgcagca 19860
 gcgggcggcc ggggtgttc gagccaccgg ctacctacc gaccaccggg ccgcacaagt 19920
 cggcaccgaa cacctggcgg cagacctgtt ggtggccttg gacgcgaacc acgctcggct 19980
 gttgcggcag ctgcgctcg aagccgcgg ggtacggatg ctgcggtcat tcgacccaag 20040
 ctcggaacc catgcgctcg atgtcagga tccctactat ggcgatcact ccgacttcga 20100
 ggaggtcttc gccgtcatcg aatccgcctt gcccggcctg cagcactggg tcgacgaacg 20160
 tcctgcgcgg aacggacgca gttgatgcc ccctagcgt tcctgtctcg gcccggtctg 20220
 ctggcgcttg cctcgtctg ggtcgcgttc acctacctgt gctttacggt gctcgcgcgg 20280
 tggcagcttg gcaagaatgc caaaacgtca cgagagaacc agcagatcag gtattccctc 20340
 gacacccgcg cggttcgcgt gaaaaacctt ctaccacagc aggattcgtc ggcgcgggac 20400
 gcgcagtggc gccgggtgac ggcaaccgga cagtacctc cggcagtgca ggtgtctggc 20460
 cgactcgcgg tgggtgaggg ggaccaggcg tttaggtgtt tggccccatt cgtggtcgac 20520
 ggcggaacca ccgtcctggt cgaccgtgga tacgtgcggc ccaggtggg ctgcacgta 20580
 ccaccgatcc ccgcctgcc ggtgcagacg gtgaccatca ccgcgggct cgtgtactcc 20640
 gaaccgagcg tggcgggcaa agaccattc gtcagagacg gcttcacgca ggtgtattcg 20700
 atcaatacgg gacaggtcgc cgcgctgacc ggagtccagc tggctgggtc ctatctgcag 20760
 ttgatcgaag accaaccogg cgggctcggc gtgctcggcg ttccgcatct agatcccggg 20820
 ccgttctgt cctatggcat ccaatggatc tcgttcggca ttctggcacc gatcggcttg 20880
 ggctattctg ctaacgcgca gatccggcgc cgcgcgggg aaaaagcggg gtcgccacca 20940
 ccggacaagc caatgacggt cgagcagaaa ctgcgtgacc gctacggcgc ccggcggtaa 21000
 accaacaatca cggccaatac cgcagcccc gccctggacca cccgcgacag caccacggcg 21060
 cggcgagat cggccacctt gggcgaccgg ccgtcgccca aggtgggccc gatctgcaac 21120
 tcatggtggt accgggtggg cccaccacgc cgcacgtcaa gcgccccagc aaacgcggcc 21180
 tcgacgacac cggcgttggg gctgggatgg cgggcggcgt cgcgcggcca ggcgcgtacc 21240
 gcaccgcggg gcgacccacc gaccaaccgc gcgcagatca ccaccagcac cgcgctgcgc 21300
 cgtgcgccaa catagtggc ccagtcaccc aatcgtgctg cagcccaacc gaatcggaga 21360
 taacgcggcg agcggtagcc gatcatcgag tcacgggtgt tgatggcacg atatccagc 21420
 acccgaggca cgcgcctcga agccgcacc agcagcggca ccacctgggc gtcggcggtg 21480
 ttttcggcca ccgactccag cgcggcacgc gtcaggcccg ggcgcgccag ctgggcgggg 21540

tcacgcccgc acagcgacgg cagcagccgt cgcgccgct cgacatcgtc gcgctccaac 21600
 aggtccgata tctggcgccg ggtgcgcgcc agcgaagttc cgcccagcgc tgcccaggtg 21660
 gccgtgcggg tggccgccac gggccaggac ctgccgggta gccgtgcag tgccgcgcgg 21720
 agcaagccca ccgcgcgcag cagcaggccg acgtgtaccg caccggcgac ccggccgtca 21780
 cggtaggtga tctgctccag cttggcgccc gcccgaccga acagggccac cggatgacct 21840
 cgtttggggg cgccgaacac gacgtcgagc aggcagccga tcagcagccc gacggccctg 21900
 gtctgcacgg tcgatgcaa cactccggca gcgtcgaca cgtggtctac gctcagctat 21960
 ttatgacctc atacggcagc tatccacgat gaagcgcca gctaccgggg ttgcgcacct 22020
 gttgaaccgg cgggcaatgt tgttcggcg agcgaatgtc atcatgcagc tggcagtgcc 22080
 ggggtgtcggg tatggcgctg tggaaagccc ggtggacagc ggcaacgtct acaagcatcc 22140
 gttcaagcgg gcccggaaca ccggcaccta cctggcggtg gcgaccatcg ggacggaate 22200
 cgaccgagcg ctgatccggg gtgocgtgga cgtcgccgac cggcaggttc ggtcgacggc 22260
 ctcgagccca gtgtctata acgcttoga ccgaagtgt cagctgtggg tggcgcggtg 22320
 totgtaccgc tactctgtg accagcacga gtttctgtac ggcccactcg aagatgccac 22380
 cgccgacccc gtctaccaag acgcaaaacg gttagggacc acgctgcagg tggcggaggg 22440
 gatgtggcgg ccggaccggg tcggttcga cagtgactgg aagcgctcgc ttgatgggct 22500
 gcagatcgac gcgcgggtgc gcgagcatct tcgcggggtg gcctcggtag cgtttctccc 22560
 gtggcggtg cgcgcggtg ccggcgctt caacctgttt gcgacgagc gatctctggc 22620
 accggagttc cgcgcgatga tgcagctgga gtggtcacag gccacgagc gtgccttoga 22680
 gtggttactt tccgtgtac ggttagccga ccggtgatt ccgcatcggg cctggtactt 22740
 cgtttaccag ctttaactgt gggacatgcg gtttcgcgcc cgacacggcc gccgaatcgt 22800
 ctgatagagc ccggccgagt gtgagcctga cagcccgaca ccggcgcggt gtgtcgctc 22860
 gccaggttca cgctcgcgga tctagagcgg ccgaaaacct acttctgggt tgccctccga 22920
 atcaacgtgc tgatctgtc gagcagctca cgcataatcg cgcgcatcgc atccacgcgg 22980
 gcatacaggt cggccttggt cgccggcagc tggtcgcagc tcattggccg caccggcggt 23040
 gctgtctgtc ggcgcgcgct gtgccttga aaccacggtc gctcacccac gaccacgaca 23100
 ctgccatata cggcgccccc ccgacaacga agcacagcta gccggtgggg cgggacggga 23160
 tcgaaccgcc gaccgctggt gtgtaaaacc agagctctac cgctgagcta cgcgcccatg 23220
 accgcgcgag gctacacgcc ttgoggccaa gcacccaaa ccttagggcg taagcgccgc 23280
 cagagcgtcg gtccacagcc gctgatcgcg aacttcaccc ggctgcttca tctcggcgaa 23340
 ccgaatgac cctgaccgat cgaccacaaa ggtgccccgg tttagcatgc cggcctgctc 23400
 gttgaagacg ccgtaggcct gactgaccgc gccgtgtggc cagaagtcgg acaacagcgg 23460

aaacgtgaat ccgctctgcg tcgccagat ctgtgagtg ggtggcgggc ccaccgaaat 23520
 cgtcagcgcg cgctgtctgt cgttctcaaa ctgggcagg tgatcagca actggtccag 23580
 ctgcacctgg cagatgcccg tgaacgcaa cggaagaac accaacagca cgttctttgc 23640
 accccggtag ccgcgcaggg tgacaagctg ctgattctgg tcgcgcaacg tgaagtccag 23700
 ggcggtggct ccgacgttca gcatcagcgc ttgccagccc gcgatttcgg ctgtaccaat 23760
 ctgctggcgc tccagttgcc cagattgacc gacgagtgct gcatcagccc agctgtgggc 23820
 gcgcctcgg caatctcgcc gggcaataca tggccgggct ggcggttctt ggcgctcacc 23880
 acccaaatca caccgtctct ggcgagcggg ccgctcgcct ccatcagggt gtccaccaaa 23940
 tcgcctgcgc catcagcca ccacaacagg acgacatcga tgacctcgtc ggtgtcttca 24000
 tcgagcaact ctccccgcga cgttctctcg atggccgcgc ggatgtcgtc gtcggtgtct 24060
 tcgtccacgc ccattctctg gataagttgg tctcgttgga tgcccaattt gcggcgctag 24120
 ttcgaggcgt gatccgcgcg gaccacgtg gaacctctt cagtctccgc gggccattgt 24180
 cacacgtgct cgatggcgat tatcgtcgca cagccagaac cggtcacacc gccgcctca 24240
 gaaggcggcc aogcacattg tcaatgcctt tgtcttgggt tcgttgagcc gatcaaccog 24300
 ccggttgaat tccgctgtcg acgcgtgcgc accgatggca ttgcccacgc ccggcgccgc 24360
 gtgcacatat cggttgagcg catccccag ttgcgcggac agcgccgcgc tcagactgcc 24420
 tgagacgcgc gaggcactgt tgttgagcgc gtcgatggcc ggaccttcgg tcggcccggt 24480
 gttgcggccc tgattgaacg cggccacgta ggcgttcaac ttgtcgatgg cgtccttgc 24540
 ggtggccgcc agcgcgctac acgaggtgcg aatcgccctt gtcgtcagcg attgttggcg 24600
 ctgcgactcc ccgatgctcg acgtcgccgc cgaagccgac accgacgcgg acaccgacga 24660
 gcggtagccc ggtgcgacgt tgggtgcggg catggccgta ccgtcggtga cagtgttaca 24720
 tccgaogac ccctacgca gcagcgcgat gcagccgagc gccagggcgc ctgcctggg 24780
 gagctccccc ccgtgcctgc gaggcacggc gcgccatcc atgagcacgg catgtgaggt 24840
 tacctggtcg cagcgcgacc gcgctggcgc tgggtgtctg cgcctccga gaaccgagcg 24900
 gagtgcggct atccgcgcgc gacgcgggtg cggcacgata gggggagcac catctaaca 24960
 gcacgcaagc ggaagccccc cacctacagg agtagtgctg tgaccaccca ttctgccgcg 25020
 cagcatctgg cccaaaactc aaacagcgca agcgaacccg accgagttcg ggtgatccgc 25080
 gagggtgtgg cgtcgtattt gcccgacatt gatcccgagg agacctcgga gtggctggag 25140
 tcctttgaca cgtctgtcga acgctgcgcg ccgtcgcggg ccgctacact gatgttgcg 25200
 ctgctagagc gggccggcga gcagcgggtg gccatcccgg cattgacgtc taccgactat 25260
 gtaaacacca tcccagccga gctggagcgg tggttccccc gcgacgaaga cgtcgaacgt 25320
 cgttatcgag cgtggatcag atggaatgcg gccatcatgg tgcacctgct gcaacgacgg 25380

ggtgtggcg tgggtggca tatctcgacc tacgcgtcgt ccgcggcgct ctatgaggtc 25440
 ggtttcaacc acttcttcog cggcaagtcg caccggggcg gcgpcgacga ggtgttcac 25500
 caggggcaacg ctccccggg aatctacgcg cgcgccttcc tgaaggggcg gttgaccgcc 25560
 gagcaactcg acggattccg ccaggaacac agccatgtcg gcggcgggtt gcgctcctat 25620
 ccgcaccgcg ggcctatgcc cgacttctcg gaattcccca ccgtgtcgat gggtttggcg 25680
 ccgctcaacg ccactaccca ggcacgggtc aaccactatc tgcattgacgc cggatcaaaa 25740
 gacacctccg atcaacaacgt gtggtgtttt ttggggcagc gcgagatgga cgaaccgcag 25800
 agcgttggcg tggcccaacgt cggcgcgctg gaaggcttgc acaacttgac ctctgtagtc 25860
 aactgcaatc tgcacgcact cgacggcccg gtgcgcggca acggcaagat catccaggag 25920
 ctggagtcgt tcttccgcg tgcggcgtcg aacgtcatca aggtgtgtgt gggccgcgaa 25980
 tgggatgccc tgctgcacgc cgaccgcgac ggtgcgctcg tgaatttaac gaatacaaca 26040
 cccgatggcg attaccagac ctataaggcc aacgacggcg gctacgtgcg tgaccacttc 26100
 ttcggccgcg accacgcgac caaggcgcgt gtggagaaca tgagcgacca ggatatctgg 26160
 aacctcaaac ggggcggcca cgattaccgc aaggtttacg ccgcctaccg gcgcgcgctc 26220
 gaccacaagg gacagccgac ggtgatccgt gccaaagcca tcaagggtca gcgcgtgggc 26280
 aagcatttgc aaggacgcaa tgccaccac cagatgaaaa aactgacctt ggaagacctt 26340
 aaggagtctc gtgacacgca gcggattccg gtcagcgacg ccagcttga agagaatccg 26400
 tacttgccgc cctactacca ccccgccctc aacgcgcccg agattcgta catgctcgac 26460
 cggcgcggcg cctcggggg ctttgttccc gagcgcagga ccaagtccaa agcgtgacc 26520
 ctgcggggtc gcgacatcta cgcgcgcgtg aaaaagggtc ctgggcacca ggaggtggcc 26580
 accaccatgg cgacggtgcg cagtttcaaa gaagtgttcg gcgacaagca gatcggggcg 26640
 cggatagtcc cgatcatctc cgacgaggcc cgcaacctcg ggttggaact ctggttcccg 26700
 tcgctaaaga tctataaccg caatggccag ctgtataccg cggttgacgc cgacctgatg 26760
 ctggcctaca aggagagcga agtcgggcag atcctgcacg agggcatcaa cgaagccggg 26820
 tcggtgggct cgttcatcgc ggcgggcacc tcgtatgcga cgcaccaaga accgatgatc 26880
 ccatttaca tcttctactc gatgttcggc ttccagcgca ccgpcgacga ctctggggcg 26940
 gcggcgacc agatggctcg aggttctgtg ctggggcca ccgcggggcg caccacctg 27000
 accggtgagg gctcgcaaca cgcgcagcgt cactogttgc tgcgtggcgc caccaccog 27060
 gcggtggtg cctacgaccc ggccttcgac tacgaaatcg cctacatcgt ggaagcgga 27120
 ctggccagga tgtgcgggga gaaccgggag aacatcttct tctacatcac cgtctacaac 27180
 gagcgtacg tgcagccgcc ggagccggag aacttcgatc ccgagggcgt gctgcggggt 27240
 atctaccgct atcacgcgcg caccagcga cgcaccaaca aggcgcgagat cctggcctcc 27300

ggggtagcga tggccgcggc gctgcgggca gcacagatgc tggccgcgca gtgggatgtc 27360
 gccgcgcagc tgttggtcgt gaccagttag gccgagctaa accgcgcagg ggtggccatc 27420
 gagaccgaga agctccgccca ccccgatcgg ccggcggggc tggcctacgt gacgagagcg 27480
 ctggagaatg ctccggggccc ggtgatcggc gtgtcggact ggatgcgcgc ggtcccgag 27540
 cagatccgac cgtgggtgcc gggcacatac ctacagttag gcaccgacgg gttcggcttt 27600
 tccgacactc gggccgcgcg tcgcgcgtac ttcaacaccg accgcgaatc ccaggtggtc 27660
 gcggttttgg aggcgttgcc gggcgacggc gagatcgacc catcggtgcc ggtcgcggcc 27720
 gccgcgcagt accggtatga gcagctggcg gctgcgcgcg agcagaccac ggatcccggt 27780
 cccggggcct aacgcggcg agccgacggc ctttgggcca atcttcacga aatctggcgt 27840
 agtttttagg agtgaacgac aatcagttag ctccagtgc ccgccgagg tcgcgcgtcg 27900
 aactgttgga cactgtgccc gattcgtgc tgcggcggtt gaagcagtac tcgggccggc 27960
 tggccaccga gccagtttcg gccatgcaag aacggttgcc gttcttcgcc gacctagaag 28020
 cgtcccagcg ccgcagcgtg gcgctggtgg tgcagacggc cgtggtcaac ttctgcgaat 28080
 ggatgcacga ccgcacaggt gacgtcggct ataccgcga ggcattcgag ctggtgcccc 28140
 aggatctgac gcgacggatc gcgctgcgcc agaccgtgga catggtgcgg gtcaccatgg 28200
 agttcttcga agaagtctg cccctgctcg cccgttcga agagcagttg accgccctca 28260
 cgggtggcat tttgaatac agccgcgacc tggcattcac ccgcgccag gccatcccg 28320
 atgcggccga ggcacgagcg acctgggaca gccggatgga ggcacgcgtg gtagacgcgg 28380
 tggtagcggc gcacaccggt ccgagctgc tgtcccgggc ggcgcgcgtg aattgggaca 28440
 ccaccgcgcc ggcgaccgta ctggtgggaa ctccggcgcc cggtcctaat ggtccaaca 28500
 gcgacggcga cagcgagcgg gccagccagg atgtccgcga caccgcggct ccgccggcc 28560
 gcgctgcgct gaccgacgtg caccgacact ggtggtggc gatcgtctcc gccacgtgt 28620
 cgccaaccga gaagtctctc aaagacctgc tggcagcatt ccgcgacgcc ccggtgggta 28680
 tcggcccccac ggcgcccatg ctgaccgcgg ccgaccgca gctagcgag gcgatctccg 28740
 ggatgaacgc cgtgcgcggc tggcgcgagg ccgcgcggcc cgtgctggct agggaaactt 28800
 tgcccgaacg ccgcctgatg ggcgacgcct ccggatcgt ggcctgcct accgacgtga 28860
 tgcggccctc agccgatgcc ggcacgacgc tcatcgagac gctagacgca tatctggatt 28920
 gtggcgcgcc gattgaagct tgtgccagaa agttgttcgt tcatccaaac acagtgcgg 28980
 accggtctaa gcggatcacc gaattcaccg ggcgcgatcc caccagcca ccgatgcct 29040
 atgtccttcg ggtggcgggc accgtgggtc aactcaacta tccgacgcgc cactgaagca 29100
 tcgacagcaa tgcgcgtgta tagattccct ccgggtcag aggggttcca gcagggggcc 29160
 cggaaagata ccagggggcg cgtcgagcgg aaagtgatcc agacaacagg tcgcgggacg 29220

atctcaaaaa catagcttac aggcccggtt tggtgggtat atacaaaaac ctaagacgag 29280
 gttcataatc tgttacaccg cgcaaaaccc tottcacagt gttctcttag acacgtgatt 29340
 gcgttgctcg caccocggaca ggggtogcaa accgaaggaa tgttgctgcc gtggcttcag 29400
 ctgcccgcg cagcgggacca gatcggcgcg tggtcgaaag ccgctgatct agatcttgcc 29460
 cggctgggca ccaccgcctc gacogaggag atccacgaca ccgcggtcgc ccagccattg 29520
 atcgtgcgc cgaactctgt ggcaccacag gaactggcgc gccgatgcgt gctcgccgcg 29580
 aaggacgtca tcgtggccgg ccaactcctc ggcgaaatcg cggcctacgc aatcgccggt 29640
 gtgatagcgc cgaagacgc cgtcgcgcgt gcgcgccccc gcggcgccga gatggccaa 29700
 gcctgcgcc cagagccgac cggcatgtct gcggtgctcg gcggcgacga gaccgaggtg 29760
 ctgagtgccc tcgagcagct cgaacttggt ccggcaaaac gcaacgcgcg cggccagatc 29820
 gtcgtgcgc gccgcgtgac cgcgttgag aagctcgcgc aagaccgcgc ggccaaggcg 29880
 cgggtgctg cactgggtgt cgcgggagcg ttccacacc agttctatgg gcccgcaatt 29940
 gacggcttg cggcgccgcg ggccaacatc gcaacgcgcg accccacgcg cagctgctg 30000
 tccaacgcg acgggaagcg ggtgacatc cggcgccgcg cgatggacac cctggtctcc 30060
 cagctcacc aaccggtgag atgggacctg tgcacgcga cgtgcgcga acacacagtc 30120
 acgcgcatg tggagtctcc ccccgcggcg acgcttagcg gtatcgccaa acgogaactt 30180
 cgggggggtc cggcacgcgc cgtcaagtca ccgcgagacc tggacgagct ggcaaaccta 30240
 taaccgcga ctcggccaga acaaccacat acccgtcagt tggatttgta cacaacatat 30300
 tacgaaggga agcatgctgt gccgtcact cagggaagaa tcattgccg tatcgccgag 30360
 atcatcgaag agttaaccg tatcgagcgc tcgagatca ccccgaggaa gtgcttgctc 30420
 gacgacctg acatcgactc gctgtcgatg gtcgagatcg ccgtgcagac cgaggacaag 30480
 tacggcgta agatcccgca cgaggacctc gccggtctgc gtaccgtcgg tgaactgttc 30540
 gccatcatc agaagctcga ggaagaaaac ccggaggcgc ctcaggcgtt gcgcgcgaag 30600
 attgagtcg agaaccocga tgcogttgcc aacgttcagg cgaggcttga ggccgagtc 30660
 aagttagtca gccctccacc gctaattggc gtttccccc cgttgtggtg accgcgctca 30720
 cagcgacgac gtcatctcgc ccggacatcg agagcactg gaagggtctg ttggccggcg 30780
 agagcggcat ccacgcactc gaagacgagt tcgtcaccaa ttgggatcta cggtcaaga 30840
 tcggcggtca cctcaaggat ccggtcgaca gccacatggg ccgactcgac atgcgacgca 30900
 tctgtactgt ccagcggatg ggcaagttgc tggcgcgaca gctatgggag tcgcgcggca 30960
 gcccgagggt cgatccgac cggttcgcgc ttgttgctcg caccggtcta ggtggagccg 31020
 agaggattgt cgagagctac gacotgatga atgcggggcg cccccggaag gtgtccccc 31080
 tggccgttca gatgatcatg ccaacgggt ccggcgcggt gatcggtctg cagcttggtg 31140

ccgcgcgcgg ggtgatgacc ccggtgtcgg cctgttcgtc gggctcggaa gcgatcgccc 31200
 acgcgtggcg tcagatcgtg atgggcgcac ccgacgtcgc cgtctcgccg ggtgtcgaag 31260
 gaccocatoga ggcgtcgccc atcgcgcgct tctccatgat gcgggccatg tcgacccgca 31320
 acgacgagcc tgagcggggc tcccgccgct tcgacaagga ccgcgcggcg ttgtgttctg 31380
 gcgagggcgg tgccgtgatg ctcatcgaga cggaggagca cgcacaagcc cgtggcgcca 31440
 agcgttgtgc ccgattgctg ggtgccgcta tcacctcgga cgcctttcat atggtggcgc 31500
 ccgcggccga ttgtgttctg gccggtaggg cgtgactcgc ctccgtggag ctggccgggt 31560
 tgtcgccggc ggacatcgac cactcaacg cgcacggcac ggccagccct atcgccgacg 31620
 ccscggaggc caaccccatc cgcgtcgccg gttgtgatca ggcgcgggtg tacgcgcga 31680
 agtgtcgcct gggccactcg atcgcgccgg tcggtgcgct cgagtcgggt ctacgggtgc 31740
 tgacgtctgc cgcagggcgc atcccccgga ccctgaacta cgagacaccc gatcccgaga 31800
 tcgacottga cgtcgtcgcc ggccaacccg gctatggcga ttaccgtac gcagtaaca 31860
 actcgttcgg gttcgccggc cacaatgtgg cgttgccct cggcggttac tgaagcacga 31920
 catcgccggg ccgcaggccc gaggtggggg tcccccgct tcggggggcg agtcggacgg 31980
 atatggaagg aacgttcgca agaccaatga cggagctggt tacggggaaa gcctttccct 32040
 acgtagtctg caccggcatc gccatgacga ccgcgctcgc gaccgacgcg gagactacgt 32100
 ggaagtgtgt gctggaccgc caaagcggga tcgtacgct cgatgaccca ttctcgtagg 32160
 agttcgacct gccagtctgc atcgccggac atctgcttga ggaattcgac caccagctga 32220
 cgcggatcga actgcgcggg atgggatacc tgcagcggat gtcacccgtg ctgagccggc 32280
 gcctgtggga aaatgccggc tcacccgagg tggacaccaa tcgattgatg gtgtccatcg 32340
 gcacccgcct gggttcgccc gaggaactgg tcttcagtta cgcagatatg cgcgtctcgg 32400
 gaatgaaggc ggtctcgcgc ctgaccgtgc agaagtacat gcccaacggg gccccgcggg 32460
 cgtcggggtt ggaacggcac gccaaaggcc ggggtgatgac gccgggtatc gcgtgcgcac 32520
 ccggcgccga gccatcgccc cgtgcgtggc agcagattgt gctgggagag gccgatgccg 32580
 ccattctcgg cggcgtggag accaggatcg aagcgtgccc catcgccggg ttgcctcaga 32640
 tgcgcactgt gatgtccacc aacaacgacg accccgcggg tgcattccgc ccattcgaca 32700
 gggacccgga cggctttgtg ttgcggcagg gcggcgccct tctgttgatc gagaccgagg 32760
 agcacgccaa ggcaactggc gccaacatcc tggcccggtt catgggcgcg agcatcacct 32820
 ccgatggcct ccacatgggt gcccccggac ccaacgggga acgcgcggcg catcgagta 32880
 cgcggcgcat tcagctcgcg ggccctcgccc ccggcgacat cgacacagtc aatgcgcacg 32940
 ccaccggcac ccaggtcggc gacctcgccc aaggcagggc catcaacaac gccttggggc 33000
 gcaaccgacc ggcggtgtac gccccaagt ctgcctcgg ccactcgggt ggcgcggctg 33060

gcgcggtcga atcgatcttg acggtgctcg cgttgcgoga tcagggtgato ccgcogacac 33120
 tgaatctcgtt aaacctcgat cccgagatcg atttgacgt ggtggcgggt gaacocgcagc 33180
 cgggcaatta cgggtatcg atcaataact cgttcggatt cggcgccac aacgtggcaa 33240
 tcgcttcgg acggtactaa accccagcgt tacgcgacag gagacctgag atgacaatca 33300
 tggccccga ggcggttggc gagtcgctcg acccccgoga tccgctgttg cggctgagca 33360
 acttcttcga cgaocgcagc gtggaattgc tgcacgagcg tgaccgctcc ggagtgtcgg 33420
 ccgcggcggg caccgtcaac ggtgtgcoga ccacgcgtt ctgcacgcag ggcaccgtga 33480
 tggcgccgcg catggcgctc gaggggtgca cgcacatcgt caacgcctac gacactgcc 33540
 tcgaagacca gagtcccatc gtgggcatct ggcattcggg tgggtcccgg ctggctgaag 33600
 gtgtgcgggc gctgcacgcg gtaggcacag tgttcgaagc catgatccgc gcgtccggct 33660
 acatcccgca gatctcgggt gtgctcgggt tcgcgcgcgg cgggcgcgcg tacggaccgg 33720
 cgttgaccga cgtcgtgctc atgggcgcgg aaagccgggt gttcgtcacc gggcccgacg 33780
 tgggtgcgag cgtcacgcgc gaggaocgtg acatggcttc gctcgggtgg ccgagagacc 33840
 accacaagaa gtcocgggggt tgccacatcg tcgcgcgaga cgaactcgat gcctacgacc 33900
 gtggcgccgg gttggtcgga ttgtttgcc agcaggggca ttctgatcgc agcaaggccg 33960
 aggcgcgtga caccgacatc cagcgcgtgc tgcgggaatc ctocgcagct gcctacgagc 34020
 tgcgtccgat cgtgacggcg atcctcgatg cggacacacc gttcgcagag ttocaggcca 34080
 attggcgccc gtogatgggt gtccggctgg gtccgctgct gggtcgcacg gtgggtgtac 34140
 tggccaacaa cccgctacgc ctggcgcggt gcctgaactc cgaaagcgca gagaaggcag 34200
 cgcgtttcgt gcggctgtgc gacgcgttcg ggattccgct ggtgggtggg gtogatgtgc 34260
 cgggctatct ccccggtgct gaccaggagt ggggtggcgt ggtgcgcgtt ggcgccaaat 34320
 tgcgtcacgc gttcggcgag tgcaccgttc ccggggtcac gctggtcacc cgaaagacct 34380
 acggcggggc atacattgag atgaactccc ggtcgttgaa ccgcgaccaag gtgttcgcct 34440
 ggccgagcgc cgaagtcgag gtgatggcg ctaaggcggc cgtcgcgcat ctgcacaaga 34500
 agaagtgtgc cgcgcgtcgc gagcacgaac gcgaagcgct gcaacgacag ttggccgcgc 34560
 agcatgagcg catocccggc ggggtcgaca gtgcgctgga catcgtgtgt gtocgacgaga 34620
 agatcgaccc ggccataact cgcagcaagc tcaccgaggc gctggcgagc gctccggcac 34680
 ggccgcggcg ccacaagaac atcccgctgt agttctgacc gcgagcagac gcagaatcgc 34740
 acgcgcgagg tccgcgcgtt gcgattctgc gtctgctcgc cagttatccc cagcgggtgc 34800
 tggtaacgc gaggcgctcc tcgcatgctc ggacgggtgc taccgacgcg ctaacaattc 34860
 tcgagaaggc cggcggggtc gccaccacgc cgaattgtct cagggtcagc acccgccaac 34920
 agctcgacgt ccaagtga aaaggcgccc tcgttcgcgt ttgttacggg gtctacgcgc 34980

cacaagagcc ggacctgttg ggccgcttgg cggctctcga tgtgttcacg ggggggcaag 35040
ccgtcgctg tctgggcacc gccgcgcgt tgtatggatt cgacacggaa aacaccctcg 35100
ctatccatat gctcgatccc ggagtaagga tgcggccac ggtcggtctg atggtccacc 35160
aacgcgtcgg tgcccggctc caacgggtgt caggctcgtc cgcgaccgag cccgcgatga 35220
ctgcgctgga ggtgcacga cagttgcgcc gcccgcgggc gctggccacc ctgcagcgcg 35280
cactacggtc aatgcgctgc gctcgacgtg aaattgaaaa cgcggttgct gagcagcgag 35340
gccgcgaggg catcgctcgg gcgcgcgaac tcttaccctt cgcgacgga cgcggggaat 35400
cggccatgga gaggcaggct cggctcgtca tgatcgacca cgggctgcg ttgcccgaa 35460
ttcaataccc gatacacgc caggttggtg aaatgtggcg agtcgacttc gcctggcccg 35520
acatgcgtct cgcggccgaa tacgaaagca tcgagtgga cgcgggacg cggagagatg 35580
tgcgcgacaa gacacgtcg gccaaagctc aagagctcgg gtggacgatt gtcccgattg 35640
tcgtcgacga tgcacagcg gaaccgggcc gcctggcgcc cgcgcatcgc cgcacacctg 35700
accgcgcgag tatggccggc tgaccgctgg tgagcagac cagagtcgca ctgcggccgg 35760
cgcagtcgca ctctcgctct gctcgcgctc aacggtgag gaactcctta gccacggcga 35820
ctaocgcctc gcgatccgtt ggcaccagac cgatccgggt ccggcggtcg aggatatcgt 35880
ccacatccag cgcctcccta tgggtcaccg cgtattcgaa ctccgcccgg gtcacgtcga 35940
tgcgctgcgc gaccgctcg gtgggcgctc cacatgtggc ggcggaagcg acgttgcccg 36000
cctcgccccc gtacgcgcgc accagcgact cgggcaatcc ggcgcccgat ccgggggccc 36060
gccacgggtt cgcgggtgcg cgatcagcg gcaggttgcg agtcggcgac ttgcggtctc 36120
gcaggttgcg cagcgtgatg gcgcgattca gcacatctc tgccatgtag cgttattccg 36180
tcagcttgcc gccgaccaca ctgatcacg ccgacggcga ttcaaaaaa cgttggtcac 36240
gcgaaacgct ggcggtgcgg cctgggacac cagcaccgcc ggtgtcgatt agcgcccgca 36300
atccgcgata ggcaccgatg acatccttgg tgccgacgc cgtccccaat cgcgtgttca 36360
ccgtatccag caggaaacgt atctcttccg aagacggttg tggcacatcg ggaatcgggc 36420
cgggtgcgtc ttctgcggtc agcccgagat agatccggcc cagctgctcg ggcattggca 36480
acaagagcgg gttcagctca ccggggatcg gaatgtcag cgcggcagtc ggaatggcaa 36540
acgacttcgc gtcgaagacc agatgtgtgc cgcggtggg gcgtagcctc agggacgggt 36600
cgatctcacc cgcaccacag ccgcgcgctg tgatgaaggc acgcgcgcac agcgcgaacg 36660
actgcggggt gcgcgggtcg gtcaactcca ccgaagtgcc ggtgacattc gacgcgccca 36720
cgtaagttag gatcgggcg ccgtgctggg ccgcggtgcg cgcgacggcc atgacacgac 36780
ggcgctcgtc gatcaattgc ccgtcgtacg cgagcagacc accgtcgagg ccgtccgcgc 36840
gaacggtggg agcaatctcc accaccggtg acgcgggat tcggcgcgat cggggcaacg 36900

tcgccgccgg cgtaccgcgt agcaccgcga aagcgtcgcc ggccaggaaa ccggcaacga 36960
 ccaacgcccg cttggtgtga cccatcgacg gcaacaaagg gaccagttgc ggcatggcat 37020
 gcacgagatg aggagcgttg cgtgtcatca ggattccgcg ttgacgggog ctgcgccggg 37080
 cgatgccac gttgcgcgtg gccagatagc gcagaccgcc gtgcaccaac ttgagctcc 37140
 agcggtcgtt googaagccc agatcatgct tttccacaaa ggccaccgto agaccgggg 37200
 tggcagcatc taaggcaatg ccaacaccgg taatgccgcc gcctatcacg atgaagtoga 37260
 gtgcgccacc gtgcggcagt gcggtcaggt cggcgaggag aocgcgcgog ttgagtgcag 37320
 ccgagtgggg catcagcaca aatatcogtt cagtgcgtgg gtaagttcgg tggccagcgc 37380
 ggcggaatcg aggatcgaat cgaagatgto cgcggactgg atggtcgact gggcgatcag 37440
 caacaccatg gtgcgccatc gacgagcgto gcgggagcgc acaactgccg accgctgcgc 37500
 cactgtcagc cggggcgcca acccctcgat caggacctgc tggtcgtgc cagagcgctc 37560
 ggtgatgtac accctggcca gctccgagtg catgaccgac atgatcagat cgtcaccccg 37620
 caaacggctg gccaccgcga caatctgctt taocaaagct tcccgctcgt ccccgctcag 37680
 gggcacctcc cgcagcacgt cggcgatatg gctggtcagc atggaagcca tgatcgacog 37740
 ggtgtccggc cagcgacggt atacggtcgg gcggctcagc cccgcgcgoc gggcgatctc 37800
 ggcaagtgtc acccggtcca cgcgtaata gcagacgcag ctgcgcgctg cccgcaggat 37860
 acgaccacog gtatccgcgc ggtcattact cattgacagc atgtgtaata ctgtaacgog 37920
 tgactcacog cgaggaaact cttccaccga tgaatggga cgcgtgggga gatcccgocg 37980
 cggccaagcc actttctgat ggcgtcgggt cgttgctgaa gcagggttgt ggcttagcgg 38040
 actcggagca gcccgaaact gaccccgccg aggtgcagct gcgcccgctc gcctgtcgg 38100
 gggcagacca 38110

<210> 25

<211> 2540

<212> DNA

<213> Homo sapiens

<400> 25

gaaaagggtg acaagtctta ttttcaagag aagatgactt ttaacagttt tgaaggatct 60
 aaaaacttgt tacctgcaga catcaataag gaagaagaat ttgtagaaga gtttaataga 120
 ttaaaaactt ttgctaattt tccaagtggg agtctctgtt cagcatcaac actggcacga 180
 gcagggtttc tttatactgg tgaaggagat accgtgcggg gctttagtgt tcatgcagct 240
 gtagatagat ggcaatatgg agactcagca gttggaagac acaggaaagt atccccaaat 300

tgcagatttta tcaacggctt ttatcttgaa aatagtccca cgcagctac aaattctggt	360
atccagaatg gtcagtacaa agttgaaaaa tatctgggaa gcagagatca ttttgcctta	420
gacaggccat ctgagacaca tgcagactat cttttgagaa ctgggcaggt tgtagatata	480
tcagacacca tatacccgag gaacctgcgc atgtattgtg aagaagctag attaaagtcc	540
tttcagaact ggcacagacta tgctcaccta accccaagag agttagcaag tgctggactc	600
tactacacag gtattgtgta ccaagtgcag tgcttttggt gtggtggaaa actgaaaaat	660
tggaacactt gtgactgtgc ctggtcagaa cacaggcgac actttcctaa ttgcttcttt	720
gttttgggcc ggaatcttaa tattcgaaat gaatctgatg ctgtgagttc tgataggaat	780
ttcccaaat caacaaatct tccaagaaat ccattccatg cagattatga agcacggatc	840
tttacttttg ggacatggat atactcagtt aacaaggagc agcttgcaag agctggattt	900
tatgctttag gtgaagtgta taaagtaaag tgctttcact gtggaggagg gctaactgat	960
tggaagccca gtgaagacct ttgggaacaa catgctaaat ggtatccagg gtgcaaatat	1020
ctgttagaac agaagggaca agaatatata aacaatatc atttaactca ttcaacttgag	1080
gagtgtctgg taagaactac tgagaaaaa ccatcactaa ctagaagaat tgatgatacc	1140
atcttccaaa atcctatggt acaagaagct atacgaatgg ggttcagttt caaggacatt	1200
aagaaaaata tggaggaaaa aattcagata tctggggaca actataaatc acttgagggt	1260
ctggttcgag atctagttaa tgctcagaaa gacagtatgc aagatgagtc aagtgcagact	1320
tcattacaga aagagattag tactgaagag cagctaaggc gctgcaaga ggagaagctt	1380
tgcaaaatct gtatggatag aaatattgct atcgtttttg ttccctgtgg acatctagtc	1440
acttgtaaac aatgtgctga agcagttgac aagtgtccca tgtgctacac agtcattact	1500
ttcaagcaaa aaatttttat gtcttaatct aactctatag taggcattgt atgtgttct	1560
tattaccctg attgaatgtg tgatgtgaac tgactttaag taatcaggat tgaattccat	1620
tagcattttg taccagtag gaaaaaaaat gtacatggca gtgttttagt tggcaatata	1680
atctttgaat ttcttgattt ttccagggtat tagctgtatt atccattttt ttactgttta	1740
tttaattgaa accatagact aagaataaga agcatcatic tataactgaa cacaatgtgt	1800
attcatagta tactgattta atttctaagt gtaagtgaat taatcatctg gattttttat	1860
tcttttcaga taggcctaac aaatggagct ttctgtatat aaatgtggag attagagtta	1920
atctcccaaa tcacataatt tgtttttgtg gaaaaaggaa taaattgttc catgctgggtg	1980
gaaagataga gattgttttt agaggtttgt tgtgtgttt taggattctg tccattttct	2040
tgtaaaggga taacacagga cgtgtgcgaa atatgtttgt aaagtgattt gccattgttg	2100
aaagcgtatt taatgataga atactatcga gccaacatgt actgacatgg aaagatgtca	2160
gagatatggt aagtgtaaaa tgcaagtggc gggacactat gtatagtctg agccagatca	2220

aagtatgtat gttgttaata tgcataagaac gagagatttg gaaagatata caccaaactg 2280
 ttaaatgttg tttctcttgg gggagggggg gattggggga ggggccccag aggggtttta 2340
 gaggggctt ttcactttcg acttttttca ttttgttctg ttoggatttt ttataagtat 2400
 gtagaccccg aagggtttta tgggaactaa catcagtaac ctaacccccg tgactatcct 2460
 gtgctcttcc tagggagctg tgttgtttcc caccaccac ccttccctct gaacaaatgc 2520
 ctgagtctg gggcactttg 2540

<210> 26

<211> 103

<212> RNA

<213> Homo sapiens

<400> 26
 agcuccuaua acaaagucu guugcuugug uuucacauuu uggauuuuccu aaauaaugu 60
 ucucuuuuuu gaaaaggugg acaaguccua uuuucaagag aag 103

<210> 27

<211> 28

<212> RNA

<213> Homo sapiens

<400> 27
 ggauuuuccu auauaaugu cucuuuuu 28

<210> 28

<211> 1619

<212> DNA

<213> Homo sapiens

<400> 28
 ccgccagatt tgaatcgcg gaccogttgg cagaggtggc ggcgggcgca tgggtgcccc 60
 gacgttgccc cctgcctggc agccctttct caaggaccac cgcctctcta cattcaagaa 120
 ctggcccttc ttggagggtc gcgcctgcac cccggagcgg atggccgagg ctggcttcat 180
 ccactgcccc actgagaacg agccagactt ggcccagtggt ttcttctgct tcaaggagct 240
 ggaaggctgg gagccagatg acgaccccat agaggaacat aaaaagcatt cgtccggttg 300

cgctttcctt tctgtcaaga agcagtttga agaattaacc cttgggtgaat ttttgaaact	360
ggacagagaa agagccaaga acaaaattgc aaaggaaacc aacaataaga agaaagaatt	420
tgaggaaact gcgaagaaag tgcgcogtgc catcgagcag ctggctgcc aaggattgag	480
cctctggcgc gagctgcoct gtcccagagt ggctgcacca cttccaggggt ttattcctgt	540
gtgccoacag ccttcoctgt ggcccocttag caatgtotta ggaaaggaga tcaacatttt	600
caaatlagat gtttcaactg tgctcoctgt ttgtcttgaa agtggaccca gaggtgcttc	660
tgcoctgtca ggggtgtctg ctggtaacag tggctgcttc tctctctctc tctctttttt	720
gggggtcoct ttttgcgtgt ttgattcccg ggcttaccag gtgagaagtg agggaggag	780
aaggcagtgt cctttttgct agagctgaca gctttgttcg cgtgggcaga gcctccaca	840
gtgaatgtgt ctggacctca tgttgttgag gctgtcacag tctgtagtgt ggaactggca	900
ggtgcoctgt gaattctgag tgcaggttcc ttatctgtca cacctgtgcc tctcagagg	960
acagtttttt tgttgttgtg tttttttgtt tttttttttt ggtagatgca tgacttgtgt	1020
gtgatgagag aatggagaca ggtccctgg ctctctact gtttaacaac atggctttct	1080
tattttgttt gaattgttaa ttcacagaat agcacaaact acaattaaaa ctaagcaca	1140
agcatttcta agtcattggg gaaaagggtt gaacttcagg tggatgagga gacagaatag	1200
agtgatagga agcgtctggc agatactct tttgcaactg ctgtgtgatt agacaggccc	1260
agtgagccgc ggggcacatg ctggcgcctc ctccctcaga aaaaggcagt ggccctaaatc	1320
ctttttaaat gacttgctc gatgtgtgtg gggactggct gggctgtgc aggcctgtgt	1380
tctgtcagcc caaccttcac atctgtcagc ttctccacac gggggagaga cgcagtcgc	1440
ccaggctccc gctttctttg gaggcagcag ctccgcagc gctgaagtct ggogtaagat	1500
gatggatttg attgcctc ctccctgtca tagagctgca ggggtgattg ttacagcttc	1560
gctggaacc tctggaggtc atctcggctg ttcctgagaa ataaaaagcc tgctatttc	1619

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/11758

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12M 1/38, 1/40; C12Q 1/68

US CL : 435/6, 91.2, 172.3, 286.1, 286.5, 282.2

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 91.2, 172.3, 286.1, 286.5, 282.2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
WEST; USPAT; DERWENT/EP ABSTRACT.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 6,060,240 A(KAMB et al.) 09 May 2000, see entire document.	1
Y	5,716,825A (HANCOCK et al.) 10 February 1998, see entire document, especially columns 7-8.	1
A	US 5,667,975 A (DYKSTRA et al.) 16 September 1997, see entire document.	1



Further documents are listed in the continuation of Box C.



See patent family annex.

*

Special categories of cited documents:

**

later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

A

document defining the general state of the art which is not considered to be of particular relevance

**X*

document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

**E*

earlier document published on or after the international filing date

**Y*

document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

L

document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

**O*

document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

O

document referring to an oral disclosure, use, exhibition or other means

**P*

document member of the same patent family

P

document published prior to the international filing date but later than the priority date claimed

Date of the actual completion of the international search

17 JUNE 2002

Date of mailing of the international search report

18 SEP 2002

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20531

Facsimile No. (703) 305-3250

Authorized officer

Valerie Bell-Harris for
BENNETT CELSA

Telephone No. (703) 308-0196